

Bytes, fast and free

In celebration of *Nature's* new website and its benefits for our personal subscribers.

Meet any science publisher, ask about electronic publishing, and — unless they know you well — witness self-restraint in action. They won't weep about copious red ink on their balance sheets. Nor will they tell you of the wee-small-hours-of-the-morning worries they have that scientists might just gain the copyright on their material. They will be tight-lipped on the obstacles to finding a perfect customizable search engine; the difficulties in turning publication into a seamless electronic interaction between authors, editors, referees and readers; the challenge of serving libraries electronically without losing subscription revenues; and the frustrating reluctance of many authors to accept that a purely electronic publication is worth anything at all.

Gradually, these problems and obstacles of perception are being overcome, while the naivety of those who believed publishing companies have no more role to play is becoming ever more clear. Moreover, although no one could describe electronic publishing as anything like mature, its huge potential for flexible and prompt delivery at affordable prices worldwide, combined with hypertext linkage between journals, is still a siren call to providers and users everywhere.

Nature's own website (www.nature.com) has been in operation since March 1996, and is visited more than 300,000 times per week. But over recent months, users will have noticed an important development: the introduction, on a test basis, of the full content of *Nature*. This issue marks the launch of the next stage in our website development: a new design, new 'web-only' content, and the availability of the full contents of *Nature* to all personal subscribers free of any additional charge. All that they have to do is to register online, supplying their subscriber number.

As well as our full printed content, subscribers will have access to web-only additional features that include in-depth software reviews,

carried out by active researchers, articles discussing the web as a medium of scientific communication, and moderated debates on topics that may be purely scientific, or which may encompass a broader scope.

But *Nature's* presence on the web is very much part of the development of the journal as a whole, rather than an adjunct to the printed version, and subscribers will benefit accordingly in a number of additional ways.

First, promptness in delivery. From 00.01 GMT every Thursday, the latest issue of *Nature* will now be available to subscribers no matter where in the world they are, or how efficient (or otherwise) their postal services. Second, interactivity: readers can adjust how they look at web pages, they can respond immediately to what they are reading, and we can quickly react to that response.

Third, the value of our scientific papers is considerably enhanced as well. Unlike any other journal, for example, direct links from references in *Nature* papers to publications across all disciplines are now available thanks to arrangements with both Medline and the Institute of Scientific Information; and very soon, NASA will be included as well. Additional content is more readily available: *Nature* has a policy of allowing authors to refer to pieces of Supplementary Information — text, audio and video — which are too large or impractical to be published in the printed journal, but which are made available by us on request. On *Nature's* website, such information will now be integrated into the electronic version of the paper.

We know that we have subscribers for whom *Nature* is not only their favourite journal, it is the best. For them, from today, the world's best scientific journal in print is also the world's best scientific journal on the web. □

Philip Campbell (*Editor*), Christopher Surridge (*Web Editor*)

Bridging economics and ecology

Economists and ecologists are increasingly working together. Such cooperation is both desirable and necessary.

What's the difference between economists and ecologists? According to one wag, economists think they are God's gift to the world, whereas ecologists think the world is God's gift to them.

This might seem to offer a poor start to a close working relationship — and, indeed, beyond this flippant answer there are many cultural differences between the two disciplines. The traditional gulf between the natural and social sciences is further widened in this case by a tendency towards opposite political leanings: the economics that is in the ascendant today is closely aligned with capitalism, yet rampant capitalism is often accused of being the enemy of the environment.

But the academic discipline of ecology is not the same as 'environmentalism', and economists are at pains to explain that the real purpose of their discipline is the efficient allocation of scarce resources — a description that is, sadly, becoming increasingly relevant to the Earth's environmental assets. There are even technical similarities between the two fields: for example, an ecologist's

description of competition and cooperation between species has much in common with the treatment of the behaviour of companies in microeconomics.

Given the crucial importance of both fields, it is not so surprising that, as described elsewhere in this issue (Briefing, page 426; Commentary, page 433), ecologists and economists are working closely together on ways to value and preserve the natural environment. Thus at one level, much of the success of the Kyoto Protocol on climate change hinges on efforts to set up an efficient and effective system of greenhouse-gas emissions trading. At another, co-operation between ecologists and economists will be needed in a \$2.3 million United Nations project announced last week to assess the demand for water in African cities, and its impact on freshwater resources and aquatic ecosystems.

In the days when humans were a small perturbation on the face of the planet, economics had little more to say about natural ecosystems than it does today about sunshine. But today, a viable combination of economics and ecology is essential; for while natural science may be



Malaria research deal seeks to make up for industry's retreat

[PARIS] The UK Department for International Development has pledged \$5 million and the World Health Organization (WHO) \$3 million for an ambitious proposal aimed at giving a major boost to research into discovering and developing new families of drugs against malaria.

The move follows the pharmaceutical industry's virtual abandonment of research on tropical diseases (see *Nature* 386, 540; 1997). The New Medicines for Malaria Venture (MMV) is a public/private sector collaboration, to be financed by public sector and philanthropic bodies, but run along the lines of an industrial R&D programme.

It marks a radical departure from current practice by acknowledging that the development of drugs for diseases largely affecting poor countries cannot be left to market forces. "If the status quo continues, the outlook for the control of many of the world's major diseases, as we approach the new millennium, looks bleak," says the proposal.

Under the proposal, academics would gain free access to industry project-management expertise and to capital-intensive technologies, such as combinatorial chemistry and high-throughput screening, that would be contributed by drug companies.

The proposal has attracted broad support. Its proponents include major funding agencies and pharmaceutical bodies, such as Glaxo-Wellcome, Hoffmann-La Roche and the International Federation of Pharmaceutical Manufacturers Association (IFPMA).

The pledges "are sufficient to get MMV off the ground," says Trevor Jones, director-general of the Association of the British Pharmaceutical Industry. They are also substantial in terms of overall funding for malaria research, which is only around \$100 million annually (see *Nature* 386, 539; 1997).

"The drug industry is saying, if you [the public sector] want us to develop drugs for diseases like malaria where there is no viable commercial market, you are going to have to spread the risk," says Simon Sargent, the representative of Glaxo-Wellcome on the MMV steering committee.

But the funds remain far short of the estimated \$30 million annual costs required to finance the complete proposal. Several funding agencies, including the World Bank, the UK Wellcome Trust and the Rockefeller Foundation — who are all members of MMV's steering committee — will want more guarantees that money will be spent efficiently before committing funds themselves.

Their reluctance stems partly from apprehension about the wisdom of funding such a

project in the public sector. It is widely recognized that the public sector lacks the expertise to take a drug from the discovery through to commercialization. "The record of drug development outside the established industry is not very good; you can spend a lot of money for little results," points out the Wellcome Trust's Bob Howells.

Under the scheme, drug-discovery research would be carried out by concentrating funds on a handful of promising projects — financed by MMV to the tune of \$3.5 million each — to put staff and resources on a par with industry-run projects. This would be carried out in partnership with companies, and is estimated at \$15 million annually.

Promising drug candidates would then be taken up to registration and phase 2 clinical trials by a 'virtual' drug development unit, financed and administered by MMV. This would require a further \$15 million annually. MMV, whose stated goal is to register a new antimalarial drug every 5 years, would seek industrial sponsors to continue development to phase 3 trials and marketing.

The criteria used to manage these operations, according to the latest draft of the proposal, would "be those of industrial manage-

ment of an R&D portfolio and not those of a public-sector science funding agency".

But many still doubt that MMV can live up to this ambition. In particular, officials from the major funding agencies are concerned that political pressure to locate MMV within the World Health Organization might prevent it from being run along business lines because, as an intergovernmental body, WHO is subject to strong political influence. "The absence of politically motivated decision-making will be key to making MMV work," says one funding agency official.

The alternative, favoured by most of the major funders, is an independent agency run by an experienced chief executive officer, headhunted from industry. Such an agency would be able to take decisions free from political influence, and to handle intellectual property and other commercial issues as it chose. Under the current proposal, the guarantee of this freedom "is still largely unclear", says Howells.

"MMV does not need to be democratic and representative, but to be led by someone with a nose for the best drugs bets who understands commercial drug development," agrees Tim Evans, from the

Titanic tourists given a 'scientific' identity

[MOSCOW] The research vessel *Academician Mstislav Keldysh*, which belongs to the cash-starved Oceanology Institute of the Russian Academy of Sciences, has provoked new controversy by apparently delivering tourists to the site of the sunken liner *Titanic* under the guise of researchers.

Deep Ocean Expeditions, a British company owned by Michael McDowell, an Australian citizen, has sold tickets at \$32,500 each to people who descend 3 kilometres in the Russian bathyscaphes *Mir-1* and *Mir-2* to see the remains of the *Titanic*.

By doing this, the company appears to have infringed a US court ruling prohibiting such trips on the basis that the exclusive rights to taking photographs of the wreck belong to the US company RMS *Titanic*.

At first, Oceanology Institute officials denied suggestion that their vessel was involved in such an activity. But a telegram sent from the ship on 15 September, and later made public, suggested that the presence of the tourists had been disguised by referring to them as scientific observers.

In the telegram, the head of the ship's 41st expedition reported that "scientific research of the first stage at the *Titanic*



In deep water? One of the *Mir* submersibles, in international, but disputed, waters.

experimental range ... is over; the observers from the United States, Germany, United Kingdom and Australia, who have invested in our expedition, took part in six dives."

Sergey Lappo, the institute's director, says they have done nothing wrong. He argues that the US court ruling applies only to US citizens, and that the *Titanic* lies in international waters.

Deep Ocean Expeditions made an advance payment to the ship to cover the 'scientific' expedition's expenses. The rest of money will go towards repairs to the ship and the two submersibles in the British port of Falmouth, to which it is now headed.

Carl Levitin

Rockefeller Foundation.

However, several academic researchers contest the plan to concentrate funding in the hands of a few groups. They argue that this could be a "lottery" that might exclude high-quality research projects, and "put all our eggs in one basket".

But Evans opposes excessive dispersion of grants, arguing that MMV amounts to a venture capital fund: "Venture capitalists are hunters who go out and see what is in the pipeline, and how these can be turned into products. What you want is a consensus as to who are the best and brightest and what the best investment portfolio would be."

Another uncertainty is the depth of industry commitment to MMV. The proposal is supported by both the IFPMA and the Hever group, an informal grouping of the research directors of the major pharmaceutical companies. At least three companies, including Glaxo-Wellcome, have also agreed to take part in the drug-discovery phase.

Sargent argues that more industry support will be forthcoming, but that it is difficult to make "tangible commitments" until discussions on specific projects are completed.

The Doldor group, an informal grouping of the chief executive officers of the major drug companies, last year rejected an earlier version of MMV. This had proposed that financing should come largely from a consortium of private companies, and the private sector still seems reluctant to commit cash.

One reason is that industry is concerned that financing such ventures directly could set a precedent, leading to pressure for similar ventures in other, more profitable therapeutic areas.

Rob Ridley, a researcher at Hoffmann-La Roche on secondment to WHO to develop MMV, is one of several observers who argue that the value of the "in kind" pledges already obtained from several companies should not be underestimated. Indeed, the MMV proposal estimates that such contributions could cut the cost of developing drugs for malaria from around \$500 million to \$186 million.

The acid test for industry's commitment will come when their support is required to take drugs from the MMV laboratory to the marketplace—the most expensive part of the drug development process. At present, this part of the MMV proposal is "very vague", says one funding agency official.

A carrot for industry being considered by the World Bank as part of MMV would involve subsidizing bulk antimalarial drug purchases by poor countries to create a market for industry. "If MMV is going to succeed, a clear commitment from the World Bank would be needed", says Sargent.

In principle, such subsidies may be relatively straightforward to arrange using existing mechanisms, according to Amie Batson, who is the World Bank's representative on MMV.

Declan Butler

Energy department under fire on pace of reforms

[WASHINGTON] The US Department of Energy (DoE) was told last week that it has not responded quickly enough to the 1995 Galvin Report, which scathingly criticized the efficiency of its huge network of research laboratories.

The report, prepared by an outside panel led by Bob Galvin, the chairman of Motorola, called for sweeping reforms to clarify the objectives of the department's laboratories and reduce their operating costs by as much as a half (see *Nature* 373, 463; 1995).

But at a hearing of the House Science Committee, the department was attacked mercilessly for its alleged sloth in implementing reforms. The criticism was based on the findings of a report commissioned by the General Accounting Office (GAO), the investigative arm of Congress.

Ken Calvert (Republican, California), chair of the energy subcommittee of the science committee, described the efforts as "baby steps where giant steps are required". Department officials "have done everything to drag their feet and not to implement even modest reforms", said Tim Roemer (Indiana), senior Democrat on the same subcommittee.

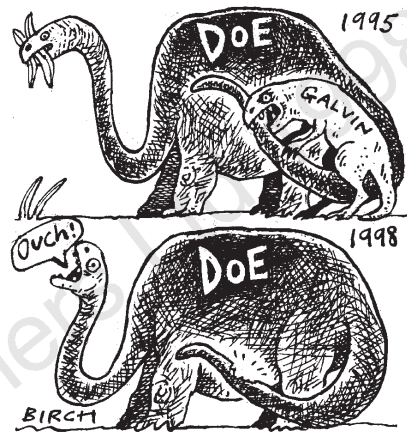
In contrast, those working at the laboratories—who have witnessed many changes in their working environment since 1995—are surprised at the GAO's conclusions. Several laboratories claim that their overhead costs have fallen by as much as a third in that time.

Ernest Moniz, the undersecretary of energy with responsibility for science and technology, told the hearing that reforms undertaken after the Galvin report would save \$2.2 billion over five years. "From our perspective there has been a lot of progress," he said. "Rome wasn't built in a day." The DoE's complex of 23 major laboratories employs 60,000 people and costs \$10 billion a year to operate.

John McTague, chief technical officer at the Ford Motor Company who co-chairs with Moniz the Laboratory Operations Board, established by the DoE to implement the reforms, told the hearing that change is underway, albeit slowly. McTague cited instability at the top of the DoE—the last secretary, Federico Peña, stayed for only 15 months—as a "major problem" in implementing change.

The GAO agrees that the Laboratory Operations Board has had a positive influence, but says the board lacks the clout to implement reforms. "It is still advisory and cannot coordinate or direct specific actions," the GAO points out.

Bruce Tarter, the director of the Lawrence Livermore National Laboratory in California, says the board has been "a very good thing" for the laboratories. "It has had a very



positive impact," he says. "The tricky part will be to maintain that."

At its most recent meeting, on 9 September, the board received a positive report from Paul Gilman of the National Research Council about the standard of peer review at the laboratories, something the Galvin report called into question.

"Merit review is pervasive in the department, and that isn't what the man on the street would think," said McTague. "There are many people in the universities who believe exactly the opposite of what's in [Gilman's] report."

Furthermore, many of the laboratories claim to have achieved sharp reductions in overhead costs in recent years. Phillip Schultz, the Lawrence Livermore laboratory controller, for example, says a 1994 initiative designed to "shine a light" on its costs has already cut overheads at the laboratory from 32.5 per cent then to 23.3 per cent this year.

Schultz's finance department has shrunk from 200 people to 140 in the process, and \$40 million a year that used to be spent on overheads has been freed up for scientific programmes there, along with another \$10 million for laboratory maintenance.

The GAO argues that most actions of this sort have been taken at the initiative of the laboratories rather than the DoE.

In contrast, Bill Madia, director of the Pacific Northwest National Laboratory, says the department "has got to be given credit" for a shake-up there which, he says, cut its overhead costs by \$60 million in just 18 months after he arrived as director.

But Victor Rezendes, director of energy and science issues at the GAO and main author of its report, is sceptical about such claims. "When you consider overheads, what they do is like squeezing a balloon," he says. "The onus now is on the Congress: there doesn't seem to be an impetus for change at DoE."

Colin Macilwain

Agencies to mitigate space station delays

[WASHINGTON & MUNICH] Life scientists and microgravity researchers face serious delays to the start of their experiments on the international space station as a result of proposed changes to its construction schedule being discussed this week in Moscow.

If US and Russian negotiators agree on a plan to shift part of the burden for supplying the station from Russian to US space vehicles, the need for extra space-shuttle missions could delay the seven scientific 'utilization' flights planned between 2000 and 2003. These would furnish the laboratory with experiment 'racks' that can be used for a variety of investigations.

Although the new assembly schedule has not been finalized, the average delay for utilization flights would probably be about six months, according to Mark Uhran of the US space agency NASA's Office of Life and Microgravity Sciences. Construction of the station starts next month, with the first utilization flight planned for April 2000.

The proposed schedule changes — made necessary by Russia's inability to pay for key elements it had promised to deliver — could also delay the completion of the station, now planned for 2004, by up to a year. Because the

laboratory module of the European Space Agency (ESA) is among the last elements to be attached, European use of the station would probably be affected.

US space managers have been struggling for months to find a politically acceptable solution to the Russian problem, which threatens the project's schedule, if not its very survival. NASA now hopes to pay Russia \$60 million immediately to finish work on a crucial 'service module' scheduled for launch next summer as the station's centrepiece during its early stages of assembly.

But a more ambitious NASA plan to buy an extra \$600 million worth of Russian hardware and services over the next four years got a chilly reception last week from key figures in the US Congress. These included House Science Committee chairman James Sensenbrenner (Republican, Wisconsin), Senate Science Committee chairman John McCain (Republican, Arizona) and House Speaker Newt Gingrich (Republican, Georgia), who called the current US–Russian agreement on the station an "absolute disaster".

With further slip in the schedule now almost inevitable, NASA hopes to soften the blow for space researchers by accommodating



Critical component: the Russian service module under construction at Khrunichev space centre.

more experiments on the proposed half-dozen extra space-shuttle resupply flights.

Another possibility is for NASA to use the Russian service module, which is nearly identical to the Mir station that housed several US experiments over the past three years. But the research community would be limited by scarce funds to flying mostly repeats of past experiments, according to Uhran.

ESA, meanwhile, wants to help its small space research community weather the delay by starting up a granting programme for ground-based experiments in life-sciences space research. The proposed scheme, suggested as part of ESA's programme for exploitation of the space station from 2000 to 2004, was discussed this week by ESA's Microgravity Programme Board, which includes representatives from ESA member states.

The scheme would give researchers access to large-scale facilities such as those designed to model aspects of living in space, including bed-rest, confinement and isolation facilities, and would help them build up facilities in their own institutes.

To complement the scheme, ESA is trying to persuade the European Commission to include ground-based space biology in its fifth Framework programme (FP5), which is due to begin operation next year, to provide researchers with project money.

"More ground-based work will allow experiments to be very carefully designed on the ground, so that we can be sure they will work in space," says Didier Schmitt, director of ESA's life sciences department.

The proposed 2000–2004 space-station exploitation programme continues microgravity work in life sciences in parallel to ground-based work, which includes sounding rocket experiments and parabolic flights.

ESA has also negotiated space for some experiments on the privately owned Space-hab, which will fly in the space shuttle's cargo bay later this month and again in 2000. "We hope the space station is not delayed," says Didier. "But if it is we hope to have a complete strategy to keep our scientific community going."

Tony Reichhardt & Alison Abbott

Japan looking forward to Christmas Island

[TOKYO] Japan's National Space Development Agency (NASDA) is to build a new space centre on Kiritimati, formerly Christmas Island, in the South Pacific, part of the Republic of Kiribati.

Programmes to be carried out there will include the development and launch of the unmanned experimental space vehicle HOPE-X, designed to ferry equipment to and from the experimental space station.

NASDA is negotiating with the Kiribati government for a 20-year lease on the southeastern part of the island. It plans to build a space port to carry out some activities transferred from its main space centre in Tanegashima, an island off Japan's south coast. These would be Japan's first space operations on foreign soil.

The Tanegashima Space Centre is shifting towards launching commercial satellites, leaving the planned space centre in Kiritimati to be

at the heart of Japan's space development programmes.

Some scientists are concerned that Japan's participation in the international commercial satellite business, scheduled to begin in 2000, will affect NASDA's research activities.

Rocket System, a Japanese company that launches commercial satellites, has signed a contract with Hughes Space and Communication International and Space Systems/Loral to launch more than 20 satellites from Tanegashima between 2000 and 2007 (see *Nature* 392, 321; 1998).

NASDA initially considered building the new centre on one of the islands in Japanese territory, but Kiritimati, near the equator, was geographically, climatically and politically more favourable. It was used for nuclear experiments by Britain and the United States between 1956 and 1963, so

NASDA is likely to face a major task in clearing up the debris that has been left behind.

There were political problems associated with most of the locations in Japan, says a NASDA spokesman, as test-flying HOPE-X would have risked entering Chinese and Korean airspace.

HOPE-X, a joint project between NASDA and the National Aerospace Laboratory, is an experimental vehicle to establish the technologies needed for the recovery and supply of equipment from the international space station.

NASDA also hopes to carry out research and development of HOPE, a space shuttle for practical missions and next-generation launch vehicles, at the planned space centre, indicating that a significant part of NASDA's R&D might be transferred to the South Pacific island. **Asako Saegusa**

NIH launches discussion of *in utero* gene therapy...

[WASHINGTON] The US National Institutes of Health (NIH) last week started what promises to be a prolonged and thorny discussion about the desirability of embarking on *in utero* gene therapy, a move that critics argue threatens to open up a "slippery slope to eugenics" (see *Nature* 395, 309; 1998).

After meeting for two days, NIH's Recombinant DNA Advisory Committee (RAC) announced a plan to set up four working groups to discuss various aspects of two newly proposed experiments, from informed consent — a clear difficulty when the subject is unborn — to issues of research design.

Following a broader 'gene therapy policy conference' on *in utero* therapy next January, the RAC will modify its guidelines, untouched since 1990, to address the issue.

The meeting focused on the dangers and merits of proposals brought to the RAC by W. French Anderson, professor of biochemistry and paediatrics at the University of Southern California, who has pioneered gene therapy in humans. Anderson and his collaborator, Esmail Zanjani of the University of Nevada School of Medicine, in Reno, are proposing two treatments for fetuses.

One treatment is for α -thalassaemia, an error of haemoglobin manufacture that in its severest form kills in the womb. Fetal blood would be withdrawn and incubated with a virus bearing the healthy form of the gene that otherwise causes production of abnormal haemoglobin. The cells would then be reinjected in the hope that they would produce normal haemoglobin.

The other treatment is for severe combined immunodeficiency (SCID), caused by a lack of the enzyme adenosine deaminase. Sufferers face immediate risk of succumbing to infection after birth. A retrovirus bearing

the normal gene for adenosine deaminase would be injected into the peritoneal cavity of the developing fetus in the hope that it will be taken up by fetal immune-system cells.

Gene therapy in the fetus is thought to increase the chance of successful gene integration because cells divide much more rapidly than in adults. But it poses various risks, from altering germ cells to introducing harmful mutations in wrongly targeted cells.

A further risk is that the treatment would bring about only a partial cure, which some meeting participants argued could be worse for children with α -thalassaemia than if they had been left to die prenatally.

There are also concerns about risks of the therapies to mothers. For instance, toxic side effects to the mother in untreated α -thalassaemia are so severe that fetuses are aborted even before they die spontaneously. Some RAC members suggested that prolonging fetal life with experimental treatment may not be justifiable.

Anderson said he brought the proposals to RAC to raise "the issue of a potential inadvertent germline gene transfer", among other risks. "If it is not possible to reduce those risks to an acceptable level [in the estimation of the Food and Drug Administration (FDA) and the NIH] we will not go forward," he said.

Claudia Mickelson, chair of RAC and the biosafety officer at the Massachusetts Institute of Technology, emphasized that the proposals were taken up for "public discussion" purposes only.

But Amy Patterson, an FDA official who this month becomes director of NIH's Office of Recombinant DNA Activities, said that, once RAC develops its policy, the FDA would need a "very compelling" reason to ignore RAC's recommendations. **Meredith Wadman**

...and seeks to repair 'flaw' in review process

[WASHINGTON] The National Institutes of Health (NIH) and the Food and Drug Administration (FDA) intend to repair a "flaw" in the system under which gene therapy experiments are approved, NIH officials announced last week.

Last year, the authority of the NIH's Recombinant DNA Advisory Committee (RAC) to approve or disapprove gene therapy experiments was shifted to the FDA. But RAC was meant to advise the

FDA on new proposals.

But RAC chair Claudia Mickelson says that protocols have been approved by the FDA before RAC, which meets only quarterly, has had a chance to comment.

Lana Skirboll, NIH's associate director for science policy, says the NIH now plans to guarantee "full public discussion of all novel gene therapy protocols before any patient is treated".

She adds that FDA and NIH are "in full agreement"

that RAC's lack of chances to provide input "is frankly a flaw in the system". Details of the new procedure are expected to be announced in January.

The Senate Appropriations Committee last month adopted language attached to a 1999 spending bill, which "strongly" encourages NIH director Harol Varmus to restore RAC's power to approve or disapprove all human gene therapy experiments. **M.W.**

Canadian research councils publish joint code on ethics

[MONTREAL] Canada's three major research fund-granting councils have published what they claim is the first broad-ranging ethics policy statement produced for research involving humans in all academic disciplines.

The statement is aimed at ensuring that research subjects will be treated with respect and privacy; that researchers and their institutions will know their work meets ethical standards; and that Canadian society will benefit from research conducted in a socially and scientifically responsible manner.

It results from several years of discussion, consultation and consensus-building among Canadian academic researchers in the humanities, social and natural sciences, medicine and engineering.

The Medical Research Council, the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council have had separate ethics policies for 20 years, but this is their first joint document. All researchers and institutions receiving their grants will have to adhere to it.

Some 350 research ethics boards in universities, hospitals and research institutes across the country review proposals for research involving humans, with authority to approve or reject proposals. The new code will update their guidelines.

A news release said the policy statement "seeks to balance the need to advance knowledge and understanding with the need to respect the existing legal, social and moral principles and responsibilities to those who participate in research as research subjects".

The document deals with consent, privacy and confidentiality; conflicts of interest; exclusion of certain groups; research involving aboriginal peoples; the conduct of clinical trials; human genetic research; and research using human gametes, embryos and fetuses.

The document cites changes in the context of research as reasons for a new approach. These included new research tools, a shift from individuals working alone to teams in centres around the world, and questions of ownership and commercialization.

Recent events have thrust such issues into the public eye. In June, for example, the pharmaceutical company Bristol-Myers Squibb attempted to stop publication of an independent report on cholesterol-lowering drugs by the Canadian Co-ordinating Office for Health Technology Assessment (CCOHTA).

CCOHTA challenged this in court, calling it an issue of freedom of speech. The company said the report was flawed and that it was not trying to stifle free speech, but to ensure that the medical profession had correct information. CCOHTA won the case. **David Spurgeon**

Report to Congress 'ducks major issues'

[WASHINGTON] The US Congress should provide "stable and substantial" funding for scientific research, with fundamental research its priority, and "resist concentrating funds in a particular area", says an 80-page science policy document prepared by Vernon Ehlers (Republican, Michigan).

The long-awaited document was prepared for Newt Gingrich (Republican, Georgia), chairman of the House of Representatives, and, according to Ehlers, is the first science policy study of its type to originate in Congress. It calls for a pilot programme to semi-privatize one of the Department of Energy's large laboratories, and for "standardized peer review procedures" at all government agencies.

But critics say the document contains little that is new, and fails to address the fundamental problem of how Congress determines science budgets. George Brown (Democrat, California), the ranking minority member of the Science Committee, says

it "undergirds the status quo" and that he will not endorse it. "It doesn't go as far or reach as deeply as I'd like," he says, although he adds that he is "willing to work" with Ehlers and others "to move forward".

Ehlers says the document — "*Unlocking Our Future: Toward A New National Science Policy*" — "demonstrates that the Congress is aware of the importance of science". He describes it as a "commencement" on what he expects will be a lengthy process to develop agreed positions on a number of science policy issues.

But Gingrich, who joined Ehlers, Brown and James Sensenbrenner (Republican, Wisconsin) for a press conference to launch the document, was notably cautious in his praise. "This is a very good start, but it really only scratches the surface of what, over the next four of five years, will have to be a very important national dialogue," he said.

Answering questions Ehlers conceded that Gingrich — who complained of scien-



Let-down? Ehlers (left) concedes Gingrich (right) was seeking more radical proposals.

tific articles "written in such a way that if you aren't in the discipline you don't understand [them]" — would have liked a more radical set of proposals, and "more risk-taking, not just with regard to science but with regard to science policy".

Gingrich asked Ehlers to carry out the study in early 1997, when Ehlers, voted second "brainiest member" in a poll of Capitol Hill staff in *Washingtonian* magazine, had failed to land the chairmanship of one of the Science Committee's subcommittees.

Scientific and engineering societies and university officials had helped Ehlers' small staff prepare the study, and most were satisfied with the result. "What's important today is having the Congress engaged," says Charles Vest, president of Massachusetts Institute of Technology and an early supporter of the project. The study, he says, "doesn't address every aspect or nuance of science policy — nevertheless it is a very thoughtful report".

But Democrat politicians say that Ehlers underestimated the difficulties of saying anything profound, and ended up leaving untouched major questions such as how to prioritize research and better coordinate the activities of different agencies and congressional committees. "It doesn't reflect the more advanced thinking about changes that need to be made — in particular, the responsibility of science to connect its work to the needs of society," says Brown.

The document is likely to be quickly endorsed by most members of the Science Committee, including some Democrats, and Ehlers hopes that a resolution in support will be passed by the House of Representatives before it goes into recess next week. Science lobbyists hope the process will help lend impetus to a proposal to double spending on research and development over the next 12 years, which currently has the support of 29 US senators (see *Nature* 394, 5; 1998). **Colin Macilwain**

US stays in global fusion deal — for a year

[WASHINGTON] Bill Richardson, the US energy secretary, signed a statement of intent last week that will allow the United States to continue work on the International Thermonuclear Experimental Reactor (ITER) design until next July.

The move has relieved fears among other nations participating in ITER that a complete US withdrawal would paralyse the global fusion energy research project (see *Nature* 394, 511; 1998).

Richardson signed the statement at the International Atomic Energy Agency conference in Vienna, after coming under heavy pressure from Japanese, European and Russian officials. Russia is said by one source to have threatened to block important agreements on the disposal of nuclear weapons materials unless the United States stayed on board.

The statement says that the United States will participate in the three-year extension to the ITER engineering design activity — which Europe, Japan and Russia agreed to in July — for a period of one year. "Participation in this process will be subject to the availability of appropriated funds and is not a commitment to construct a device," it says, reiterating the US desire for "a new international agreement on fusion science".

Richardson called Joseph McDade (Republican, Pennsylvania), chairman of the House appropriations subcommittee which funds his department, from Vienna to obtain his agreement to the statement. But both the agreement and the call to McDade infuriated another congressman, James

Sensenbrenner, the chairman of the House Science Committee, who was not consulted.

Sensenbrenner immediately issued an angry statement condemning the deal, and describing the decision as "irresponsible". He added: "The project has failed and it's time to move forward. It defies common sense that the United States should agree to continue to participate in a dead-end project that continues to waste the American taxpayer's dollars".

Energy department officials were surprised at the vehemence of this, pointing out that Sensenbrenner had previously said that he was neutral on ITER, and merely wished McDade's concerns to be addressed. Ernie Moniz, the energy undersecretary, wrote to Sensenbrenner on his return from Vienna saying that the twelve month agreement will "allow for an orderly closing out of US participation under the existing agreement, and a reasonable transition to a new one."

Even at that, the statement was greeted with relief in the US fusion community. "I'm very, very pleased," says Anne Davies, head of fusion energy sciences at the Department of Energy. "This gives us a way forward."

Susumu Nakamura, director of fusion energy at Japan's Science and Technology Agency, says: "It is encouraging that the United States is at least showing some willingness to continue with the ITER collaboration". But he also says that the STA "believes that the one year extension is renewable," a view which Moniz's letter appears to contradict. **Colin Macilwain**

UK university fund 'seeking out the best'

WELLCOME TRUST

[LONDON] High-quality schemes that are both ambitious and imaginative are being sought for a new £600-million fund for research infrastructure being jointly funded by the British government and the Wellcome Trust.

This is the message that both the government and the trust say they want to convey to Britain's scientific community as they open the first call for applications today.

Mike Dexter, director of the Wellcome Trust, argues that a shortage of funds has forced universities to adopt a conservative approach to applying for support. "We want to change this culture of undercutting, where universities write out applications on the basis that the lowest-value bid will get funded."

Richard Lane, programme director of the fund at the Wellcome Trust, adds: "We want the best science, but not just safe science. In the long-run, cheap does not necessarily mean value for money."

Dexter says the fund is looking for imaginative, perhaps cross-disciplinary, ideas that may need time to develop and may not lead to an immediate flow of research publications.

The government and the trust will each contribute £300 million to the fund, which was announced in the July budget (see *Nature* 394, 209; 1998) and will last until 2001. It is designed primarily to purchase equipment and for the construction and refurbishment



Dexter: in quest of imaginative ideas.

of laboratories. Bids below £750,000 are unlikely to be considered.

Government officials and trust representatives will embark on a 'road-show' of six universities in England, Wales, Scotland and Northern Ireland on 9 October, introducing the fund and giving potential applicants a

feel for types of projects likely to be funded.

Sir John Cadogan, director general of the research councils, says universities would be wrong to see the fund merely as a way of fixing old equipment, as applications had to be accompanied by plans for "high-level science".

Cadogan will chair the fund's joint executive committee, the body that will ultimately decide which projects to support. In January he will hand over to his successor, John Taylor, director of Hewlett Packard's European research laboratories in Bristol. Dexter will be the committee's deputy chair.

The committee is made up of government officials, the heads of the research councils, representatives of the Wellcome Trust, and the heads of the four UK higher-education funding councils, who are non-voting members.

Both Cadogan and Dexter believe that more than half of the fund is likely to be allo-

cated to applications with a biological science component, largely because of the trust's legal status as a charity devoted to biomedical research. But they say that there are no quotas for any particular type of research, nor will the fund be divided according to regions.

Lane and Dexter emphasize that the fund will not "in any way" reflect the trust's own research priorities, which are taken care of under separate programmes. It remains unclear at present the extent to which proposals selected for funding will be associated with topics identified as potential priorities in the government's Technology Foresight exercise.

An international scientific advisory board will assess all applications with an emphasis on the biomedical and biological sciences, including environmental science. The board will be chaired by Richard Flavell, professor of immunobiology at the Yale University School of Medicine. Its members have yet to be confirmed, but two-thirds are expected to come from outside the United Kingdom.

Existing peer review panels from four research councils — Engineering and Physical Sciences, Particle Physics and Astronomy, Natural Environment, and Economic and Social Research — will assess all other research applications. The international scientific advisory board and the joint executive committee will hold five application rounds between now and 2001.

Ehsan Masood

French government tightens its grip on research priorities

[PARIS] France has decided to create a centralized FF500 million (\$89 million) annual fund for basic research within the Ministry of Research. Although some of the money for the so-called National Science Funds (FNS) has been taken from existing ministry budget lines, much of it is new.

The fund forms part of the French civil budget for research and development for 1999, the distribution of which indicates a shift towards greater ministerial control and a preference for the universities over the public research agencies, such as the Centre National de la Recherche Scientifique (CNRS).

The overall budget is expected to increase slightly, by 1.6 per cent, to FF53.9 billion. The new fund will be used mainly to finance multidisciplinary research projects, emerging areas of research, and in particular programmes that span several research organizations. The latter indicates a shift to carrying out strategic research, such as genomics, at a supra-agency level.

Strategic research directions will be decided upon by a new 28-member advisory committee, the National Science Council, chaired by Claude Allègre, the minister for national education, research and technology.

The committee, a third of whose members will be from overseas, will be attached to the prime minister's office.

It is not yet clear how grants will be selected. But according to Vincent Courtillot, Allègre's principal adviser, the process will rely on an open call for proposals and external peer review administered by a series of new ministerial committees, such as that recently created to coordinate life sciences (see *Nature* 395, 315; 1998).

Management of cross-agency projects would be delegated to the research organization with most expertise in the field concerned. This may ease trade union fears that grant distribution within programmes may be centralized within a few ministerial committees rather than being carried out by the research organizations themselves.

Courtillot says the fund represents a 'top-down' effort by the ministry to fix the broad directions of strategic research areas, while maintaining an investigator-driven 'bottom up' approach to distributing grants within these areas.

The Funds for Technology Research, which had accumulated a debt of FF1 billion, is back in the black and will receive FF660

million in the new budget. Priorities include nanotechnology, drug development and new materials. A further FF200 million seed fund has been created to encourage start-up technology companies.

A breakdown of the budget allocations for individual research organizations has not been finalized, but according to Courtillot these will increase by 2.2 per cent overall, with funding for laboratory equipment and supplies being increased by 8 per cent, by making economies in administration and shifting programme research to the national level.

CNRS declines to comment on what these figures will mean on the ground. But Audier's analysis of the figures suggests that the CNRS budget is at best "very mediocre".

The universities' research budget will rise by 2.9 per cent, with fundamental research increasing by 7.3 per cent, according to the ministry. The latter is a "political message" that the university is the natural home for research, says Courtillot. Similarly, while 100 research posts and 50 technician posts will be created within the research organizations, the universities will obtain 1,500 assistant lecturer posts.

Declan Butler

Australian opposition pledges more funds

[SYDNEY] In a tightly fought election campaign in Australia, Kim Beazley, the leader of the opposition Labor Party, has proposed science and higher-education research initiatives costing about A\$600 million (US\$355 million) over three years, saying that this will lead to lower unemployment.

In contrast, the Liberal Party's science and education ministers have been low key and defended cuts by the conservative coalition, of which it has most members, to both sectors over its two-and-a-half years in office. But the Labor Party would need a large swing to regain power in the election, which takes place on Saturday (3 October).

Only some of Labor's proposed funding for science would need 'new money'. One bold suggestion is that 35 per cent of the profits of Telstra, Australia's largest company, should be spent on science and development projects outside the main cities.

Last year, the government sold off one-third of Telstra, a telecommunications company. John Howard, the prime minister, now wants to sell the remaining two-thirds. But Labor and minor parties are vigorously opposed to further privatization. Labor says that, if elected, it would set up a Telstra Reward Fund, the annual income of which would be the equivalent of the dividends earned by international shareholders if Telstra were fully sold.

Labor estimates that the fund's income in the first year would be sufficient to pay for grants awarded through the National Health and Medical Research Council, the 67 cooperative research centres and the Australian Institute of Marine Science, and pay for a Molecular Biology Research Institute, creating 350 jobs in Adelaide.

Labor's shadow science minister, Martyn Evans, says the Telstra plan would "isolate science from the normal slings and arrows of outrageous fortune of the budget process". He also argues that "flagging" support for science, from a public service heavily dependent on innovation, could increase public appreciation of science.

Beazley says that Labor's policies would raise business spending on research and development — currently falling fast — to more than one per cent of gross domestic product by 2005–2010, "restoring Australia's reputation as a clever country".

Evans says a A\$300 million research reactor — Australia's largest investment in a science facility — would not be built on a site in Sydney chosen by the ruling coalition, and the necessity for one anywhere else would be reviewed. The minority Democrats Party, which is likely to hold a balance of power in the Senate, is totally opposed to it.

Brian Anderson, the president of the Academy of Science, says he is "alarmed and

dismayed" at recent evidence of a fall in industrial research and development (see *Nature* 394, 512; 1998), and describes Labor's promise to restore tax concessions as "a starting point for constructing a taxation regime for innovation that will work".

In a costly advertising campaign fronted by Sir Gustav Nossal, previously director of the Walter and Eliza Hall Institute for Medical Research in Melbourne, university vice-chancellors have called for about A\$1 billion cut from universities' budgets to be restored.

The vice-chancellors have greeted Labor's proposals — which include the reversal of a substantial reduction to the Australian Research Council foreshadowed in the coalition's budget last May — as "the first breath of fresh air in a long time".

John Moore, the minister for industry and science in the current government, has promised increased funding for a new scheme that provides competitive grants to industry, and modest initiatives in marine

science and wetlands research. But Anderson says the country's problems are "only lightly touched" by such suggestions. **Peter Pockley**



Canberra's Telstra Tower: Labor wants Telstra's image — and profits — to boost science.

JACOB LILDBALLE

NIH ponders role of 'public' advisers

[WASHINGTON] Harold Varmus, the director of the National Institutes of Health (NIH), last week turned to 25 individuals — from an Illinois farmer's wife to a former African ambassador — for advice on establishing a Council of Public Representatives.

The move represents Varmus's swift response to a report published three months ago by the Institute of Medicine (IOM) at the request of Congress that criticized the NIH for failing to communicate effectively with the public (see *Nature* 394, 111; 1998).

The council, which Varmus hopes to have in place within six months, was recommended by the IOM report as a way of improving relations between the NIH and the public by getting ordinary people's opinions directly to Varmus.

Addressing a public meeting of the 25 individuals, Varmus added that he has already implemented another of the report's recommendations by establishing public liaison offices in all 24 NIH institutes and centres. And he is also acting on another IOM suggestion: adding public members to the director's advisory committee.

While Varmus stressed that the IOM had issued "recommendations, not orders", he said that the NIH "can always do better" with public relations, and he sees a valuable role for the new council. "It would be very helpful to me as we go through debates about issues where science touches society that I have some standing body that has more public participation, that can give me advice," he said.

In the course of the day-long meeting,

however, it became clear that defining the role and composition of the body will not be simple. Some participants wanted the group to have a clear hand in advising Varmus on research priorities, the larger subject of the IOM report.

Others, however, argued that it would not be scientifically equipped for such a role, and would be more effective focusing on broader issues, such as the public's concerns about medical privacy.

Rashi Fein, a professor of the economics of medicine at Harvard Medical School, suggested that the group should focus on "certain gaps in the sensitivity of the medical science community". But Amalie Ramirez, associate director of the Baylor College of Medicine Center for Cancer Control Research in San Antonio, Texas, was worried that the council might then be relegated to a "public relations function".

Still others argued that neither the council nor the public liaison offices address what they perceive as the problem underlying NIH's public relations efforts: the agency's apparent blindness to perceived imbalances in how it distributes its \$14 billion annual budget among diseases.

But Varmus cautioned against a council structured in such a way that meetings became "a divisive debate among constituencies wanting bigger shares of the pie". He called for the group to help develop membership criteria that would offer him "protection" from the demands of various advocacy groups to be included in the council.

Meredith Wadman

US rally calls for doubling of funds to cancer institute

[WASHINGTON] An estimated 150,000 people converged on the heart of Washington DC on Saturday (26 September) to demand more and better cancer research, including a doubling in funding for the National Cancer Institute. They heard speeches from Vice-President Al Gore, whose sister died of lung cancer, and Queen Noor of Jordan, whose husband, King Hussein, has cancer.

On the same day, President Bill Clinton devoted his weekly radio address to cancer, announcing that by next year all federal cancer research programmes will “fully integrate patients and advocates into the process of setting research priorities”. Congress itself is poised to grant a generous increase in cancer funding, which the administration has suggested should rise by 65 per cent over the next five years, in the coming fiscal year.

But the bill containing the funding — along with that for the rest of the National Institutes of Health — is currently stalled by Capitol Hill politics. It may be included in an ‘omnibus’ spending bill funding three to five other

departments and agencies that Congress aims to pass before adjourning later in the month.

Brussels roasts Portugal over BSE in cattle

[PARIS] Concerns that some European countries are not doing enough to counter the risk of an increase in the incidence of bovine spongiform encephalopathy (BSE) in cattle (see *Nature* 395, 205; 1998) took a political turn last week when the European Commission issued a stern warning to Portugal to improve controls or face sanctions, including a possible beef embargo.

The 60 BSE cases declared in Portugal are already more than the total for all of 1997 — between 1994 and last year the country declared 147 cases. The warning, by Franz Fischler, the agriculture commissioner, follows an inspection in Portugal by commission veterinary inspectors, who claim shortcomings, for example, in the enforcement of a ban on feeding meat and bone meal to cattle.

Cancer researchers win Lasker awards

[LONDON] This year’s Albert Lasker awards have gone to six cancer researchers. Lee

Hartwell, president and director of the Fred Hutchinson Cancer Research Center in Seattle, Yoshio Masui, professor emeritus of zoology at the University of Toronto, and Paul Nurse, director-general of the Imperial Cancer Research Fund in London, have been honoured for basic research into the machinery that controls cell division.

Alfred G. Knudson Jr, former president of the Fox Chase Cancer Center in Philadelphia, Peter C. Nowell, professor of pathology and laboratory medicine at the University of Pennsylvania School of Medicine, and Janet D. Rowley, professor of medicine, molecular genetics and cell biology at the University of Chicago Medical Center, received their awards for work into the genetics of cancer. Daniel E. Koshland Jr, former editor of *Science*, received a special award.

Cornell investigates claims of falsified data

[WASHINGTON] Cornell University has opened an investigation into allegations that an associate professor of medicine and microbiology at the university’s Joan and Sanford Weill Medical College and Graduate School of Medical Sciences ordered the falsification of data in a research grant application, knowingly used false claims to win another \$1.5 million federal grant, and

published a paper based on falsified experiments.

Laboratory workers have also alleged that immunologist John Ho threatened or punished them when they confronted him with evidence of misdeeds. Ho told *The New York Times* that he did not believe he had intentionally conducted scientific misconduct. The experiments under scrutiny involve the interaction between the immune system and microorganisms, including HIV and tuberculosis.

Congress agrees to more high-tech visas

[WASHINGTON] The US House of Representatives last week passed legislation raising the number of non-US high-tech workers allowed into the country. The quota for H-1B visas, which includes university researchers, computer programmers and other employment categories, is set at 115,000 for the fiscal year beginning 1 October, an increase of 50,000 over last year's allowance (see *Nature* 394, 610; 1998).

The quota would remain at 115,000 in 2000, fall to 107,500 in 2001, and then revert to 65,000 in 2002. A compromise reached last week between congressional advocates of higher quotas and the White House, which had threatened to veto the legislation unless

it included stronger protection for US workers, makes it virtually certain the Senate will pass the bill.

UK warning over fall in university technicians

[LONDON] The number of technicians at UK universities has fallen sharply over the past 15 years, according to a report from the Royal Society published last week. The report says that any further reduction in their number "would have a severely detrimental effect upon the quality of UK scientific research". The report calls on funding councils and charitable bodies to find ways of maintaining a "core" level of scientific support staff.

The number of university-funded technicians has fallen from 18,239 in 1982 to 12,301 in 1994. At the same time, the number of full-time, university-funded academic staff has remained broadly constant at around 31,000, while the number of externally funded researchers has increased from 11,000 to 18,000.

India now ready to sign nuclear test ban treaty

[NEW DELHI] Defence scientist A. P. J. Abdul Kalam and Atomic Energy Commission chairman Rajagopalan Chidambaram, the

key individuals behind the country's recent nuclear tests (see *Nature* 393, 197–198; 1998), have announced that India is now in a position to sign the Comprehensive Test Ban Treaty (CTBT) as no further underground tests are needed.

Kalam told a press conference last week that the five tests carried out during May, and the one conducted in 1974, have provided sufficient data for computer simulation for weapons design and to ensure that only subcritical tests not banned by the CTBT are needed in future.

Russia appoints new science minister

[MOSCOW] Mikhail Kirpichnikov has been appointed minister of science and technology in the new cabinet headed by prime minister Evgeny Primakov. He was previously a deputy to the former minister, Vladimir Bulgak, who has been promoted to vice prime minister.

Kirpichnikov is a biologist who was elected to the Russian Academy of Sciences while head of the government's science and education department. The move was a controversial one at the time, and many scientists argue that it creates a potential conflict of interest, as the minister is expected to oversee the fair distribution of research funds across the whole scientific community.



Fetching water in Malaysia: should — and can — economists take account of the rights of future generations to a clean environment?

Costing the Earth: when ecology meets economics

In recent years, ecologists and economists have learned the importance of working together. But the alliance has at times been rocky, particularly as there is no consensus on how cooperation is most effectively achieved.

Almost 50 years ago, British scientists identified several thousand sites containing rare plants, animals and geological features. The government ordered their preservation, and their owners were compensated for not using them, the amount of compensation being based on how much profit landowners would lose, for example, by not planting crops.

Collaborative research between economists and ecologists is thus a well established route to tackling environmental issues such as biodiversity conservation, climate change, and the management of forests.

Until recently, economists and ecologists, when collaborating, have tended to work from within their own disciplines, but with an unwritten hierarchy between the two:

This briefing has been written by Ehsan Masood in London and Laura Garwin in Washington

those known as 'environmental and resource' economists have tended to frame the problem, with ecologists providing the basic data.

This hierarchy, however, is now being challenged by two distinct schools of a newer discipline known as 'ecological economics'. Broadly speaking, members of both schools no longer accept ecology as merely an input into environmental and resource economics. They want trained ecologists playing a greater role in setting — and not just achieving — conservation research goals. Both schools are also united by a conviction that placing a financial value on environmental services is one way of focusing attention on the importance of environmental protection.

But that is where the similarities end. One school wants to work within the parameters of neoclassical economics. The second refuses to, and, ambitiously, wants ecological economics seen as an entirely new discipline.

Environmental and resource economics was formally established in the United States in 1975 with the setting up of the Association of Environmental and Resource Economists, with strong backing from a Washington-based think-tank, Resources for the Future. The association now has 850 members, as well as a journal, *The Journal of Environmental Economics and Management*.

Initially, "we were the rebels from mainstream economics", says Nancy Bockstael, professor of agricultural and resource economics at the University of Maryland at College Park, and president-elect of the Association of Environmental and Resource Economists. "But we are now viewed [by neoclassical economists] as a legitimate field."

A political alternative

The origins of ecological economics, which emerged as a discipline in its own right in the mid-1980s, are less clear. But there is general agreement that it resulted partly from a feeling that environmental and resource economics was getting needlessly mired in theoretical issues. "I'm just tired of the questions that turn on environmental and resource economists," says Stephen Farber, an economist at the University of Pittsburgh. "They'll be differentiating equations while the world collapses."

Another factor was more political, namely a bid by ecologists for a greater role in shaping the environment/economics research agenda alongside economists, not, as one observer notes, as their "contract researchers".

On other pages | [Could 'enviro-capitalism' be the answer?: p428](#) | [What price a greener GDP? p428](#) |

Issues that ecological economists wanted to research included comparing methodologies for environmental impact assessments, and costing ecosystem 'services' such as the pollination of plants by insects. Other goals have included factoring depletion of natural resources, such as gas and petroleum reserves, as well as the costs of ecosystem services, into national accounting systems. Some have also been involved in developing market-style incentives to help conserve natural ecosystems (see page 428).

Schools of thought

One school of ecological economics has strong links to the Beijer Institute of Ecological Economics in Stockholm, set up by the Royal Swedish Academy of Sciences in 1991.

The Beijer Institute is governed by an equal number of ecologists and economists. It has a journal, *Environment and Development Economics*. Its affiliates include prominent figures from mainstream ecology and economics, including Paul Ehrlich, Bing Professor of Population Studies at Stanford University, and Nobel laureate economist Kenneth Arrow, professor emeritus at Stanford.

The second school of ecological economics revolves around the 1,500-member International Society for Ecological Economics (ISEE). The ISEE also has a peer-reviewed journal, *Ecological Economics*. But its members have more diverse backgrounds than those attached to the Beijer Institute. Around two-thirds are ecologists or economists, with the rest coming from other fields; a few are ex-Marxist economists.

The ISEE was founded by Herman Daly, a prominent economist and former World Bank official, now at the University of Maryland. Daly is known, among other things, for propagating the view that there is a limit to how long the Earth's natural resources can be consumed at the present rate — and that this limit is not far off. Daly believes that ways need to be found to reduce human consumption, as measured by the rate of economic growth, but without reducing quality of life.

Daly's influence is one reason why many of the ISEE's members are inclined to question neoclassical economics. Its members do not share the optimism of neoclassical economists that a combination of technology, regulation and market mechanisms will be sufficient to solve environmental problems such as energy and food shortages.

"We get all the stakeholders together — ecologists, economists, wildlife managers, local people — and say: 'here's some resource we're interested in, how do we understand it, and how do we manage it?'" says Robert Costanza, who is the director of the University of Maryland Institute for Ecological Economics and another of the founders of the ISEE. "The result is a model that isn't an

ecological model or an economics model; it's an integrated model."

One example of an ISEE ecological economics approach to a problem is given by Richard Norgaard, of the University of California, Berkeley, who is the ISEE's incoming president. He points out that, in assessing industrial pollution, a neoclassical economist would subtract the 'costs' of the pollution (for example in terms of climate change or damage to human health) from the 'benefits' measured in terms of industrial output and employment.

Based on this analysis, he says, a neoclassical economist would calculate what was considered to be a sustainable level of pollution, usually a quantity whose benefits are greater than the costs. In contrast, says Norgaard, an ecological economist will factor in additional elements, such as the right of future generations to a clean environment. This will result in a level of acceptable pollution lower than that arrived at by neoclassical economists.

Perhaps a more controversial example is a recent study that attempted to place a single monetary value on all of the Earth's ecosystem services (see *Nature* 387, 253–260; 1997). This study, according to Costanza, its lead author, consciously chose to use concepts from economics. However, according to neoclassical economists, it disregarded their definitions of how these concepts should be used (see page 430).

Norgaard acknowledges that his vision of ecological economics is not yet sufficiently developed to offer prescriptions for effectively managing environmental resources, or for shrinking economies without affecting quality of life. But he believes that, in time, these will come.

Disagreement over methods

What do members of the Beijer school of ecological economics make of all this? Some may agree with the second school's overall aims, but most are unconvinced of the methods.

Karl Göran Mäler, director of the Beijer Institute, is one such critic, choosing not to invite prominent ISEE members — including Costanza — to a meeting at the institute last month intended to produce a consensus on valuing the environment.

Another critic is David Simpson, an environmental and resource economist at Resources for the Future, but who also has links to the Beijer Institute. Simpson, who specializes in valuing biodiversity, argues that in practice most policymakers are concerned more with practical ways of managing resources sustainably than with developing new approaches to environmental problems.

He also believes there is an almost universal political consensus that economies need to grow — albeit sustainably — in order to prosper. And he sees much "reinventing of

the wheel" when watching ecological economists — from both camps — struggle to redefine well-established concepts in economics. "At the risk of sounding pedantic, I think it is important to have respect for each other's area of study," he says.

Partha Dasgupta, professor of economics at the University of Cambridge and a past chairman of the Beijer Institute, thinks that many of the differences among all three strands — the two schools of ecological economics, and environmental and resource economics — are largely down to differences in rhetoric rather than substance. Indeed, Costanza is on Beijer's visiting staff.

For example, Dasgupta says that when it comes to solving actual problems, environmental and resource economists and ecological economists (of the Beijer school) both work within the parameters of neoclassical economics.

But, while the two schools of ecological economics may share common goals and ways of working, there remains a fundamental difference which is difficult to ignore.

Ecological economics, according to a definition proposed by Daly, Costanza and Norgaard, aims to provide a framework for the equitable distribution of resources and property rights within the present generation of humans, between current and future generations and between humans and other species.

To some, this sounds more like a political programme than the definition of a standard academic discipline, and suggests that ecological economics — of one variety at least — has a clear social agenda. It may be the willingness or otherwise of economists to take this agenda into account in their work on environmental issues that represents the sharpest division between different schools.



Going down: can a price be placed on the cost of depleting the world's oil reserves?

When self-interest is key to a better environment

As economists and ecologists seek to build intellectual bridges between their disciplines, economic incentives are already being used to protect the natural environment and the 'services' it provides to society. These incentives take many forms, from the tourist dollar to tradable pollution permits. All are becoming more effective as natural amenities become scarce — and therefore more highly prized.

Perhaps the most straightforward example of 'doing well by doing good' is 'ecotourism'. For example, the South African company ConsCorp (Conservation Corporation) has agreed with local landowners to restore several hundred thousand hectares of farmland to their original state and to stock the land with animals. Land that yielded \$25 to \$70 per hectare a year for ranching or farming now yields \$200 to \$300. Visitors pay a premium to see (and hunt) lions and leopards, so there is a great incentive to maintain the ecosystem needed to support these pinacles of the food chain.

Clean water cheaply

Conservation can also lead to the avoidance of costs that would otherwise be incurred. For example, protecting watersheds from development is a relatively cheap way to provide clean, abundant water for downstream users.

New York City avoided paying more than \$6 billion for a water filtration plant (plus running costs of about \$300 million a year) by investing \$1–\$1.5 billion in restoring the soil ecosystems of the Catskill mountains watershed (see *Nature* 391, 629–630; 1998). By buying land in the Catskills and restricting its use, the city government was able to preserve the water filtration capabilities of the watershed, instead of replacing this 'service' with a more expensive, engineered solution.

Extended globally, there could be an economic justification for conserving up to 13 per cent of the world's land area for its watershed value to urban water supplies, according to Walter Reid, a visiting fellow at the World Resources Institute in Washington DC.

In these examples, an economic incentive exists for preserving an ecosystem because it supports goods or services with a clear market value. These goods and services have the property of 'excludability' — it is possible to exclude people from consuming them, and so make people pay for such consumption.

Most of the services provided to humans by ecosystems, however, have a non-excludable character; that is, they provide benefits to people who may never set foot in the ecosystem, or even be aware of its existence. Such

services include the retention of soil and prevention of flooding by vegetated landscapes, and the role of birds and insects in pest control and pollination. In the most extreme example of a shared benefit, the sequestration of carbon by a hectare of forest benefits all humans by offsetting greenhouse-gas emissions. The owner of land supplying such services does not receive any compensation for them, and so has no economic incentive to conserve the resource.

The economist's solution to the problem of non-excludability takes the form of assigning an appropriate kind of property rights, in the same way that patents and copyrights protect knowledge. In the environmental realm, examples include grazing rights, fishing quotas and water rights.

For maximum economic efficiency, these rights should be 'tradable', giving them a market value and creating incentives for conservation. The US government has addressed many of its air pollution problems by means of emissions permits. One notable success has been the programme of tradable sulphur dioxide emissions allowances set up by the

Clean Air Act amendments of 1990. In this programme, the targeted emissions reductions have been exceeded, saving \$1 billion a year compared with non-market-based 'command and control' policy alternatives.

Largely as a result of pressure from the United States, the Kyoto Protocol on climate change goes further, providing not just for tradable carbon dioxide emission permits, but for permits to be given to countries that contrive to sequester an equivalent amount of carbon — for example, by reforestation.

This could profoundly affect the economics of forest conservation, especially as the protocol's Clean Development Mechanism would allow developed countries to pay for forest conservation in developing countries. Geoffrey Heal, of Columbia University's Graduate School of Business, estimates that growing forests might earn carbon sequestration credits at a rate of \$70 to \$800 per hectare per year. This compares favourably with maximum annual profits from ranching in Costa Rica of \$100 to \$125 per hectare.

Not everyone is enthusiastic about tradable emissions permits, which some environmentalists characterize as "licences to pollute". Robert Costanza of the University of Maryland prefers pollution taxes, on the grounds that these give polluters an incentive to reduce emissions "all the way down to

Progress and pitfalls along the path towards

One goal that unites most ecological economists is the desire to develop a system of national accounting that embraces environmental factors excluded from current definitions of gross domestic product (GDP).

The conventional assessment of GDP, which is around 50 years old, works by adding up all the final demands for goods and services produced annually by a nation. Although widely used by economists, journalists and politicians as the measure of the economic health of a country, GDP has been much criticized by environmentalist groups — backed by some sympathetic economists — on the grounds that it paints a potentially misleading picture of a society's health when seen in environmental terms.

"A country could exhaust its mineral resources, cut down its forests, erode its soils, pollute its aquifers, and hunt its wildlife and fisheries to extinction, but its GDP would not be affected as these assets disappeared," says Robert Repetto, an environmental and resource economist at the World Resources Institute, an independent research organization in Washington DC.

Repetto has pioneered work on a greener GDP. His study in Indonesia in 1989 concluded that annual GDP growth

corrected for depreciation in timber, petroleum and soil resources was 3 per cent lower than the conventionally calculated figure of 7.1 per cent between 1971 and 1984.

More radical ecological economists such as Herman Daly (see page 427) take a different view. They criticize the idea that a country's wealth can be measured just in terms of how much its citizens produce.

Daly has helped to develop what he and colleagues refer to as the Index of Sustainable Economic Welfare (ISEW). Although this takes GDP as its starting

How green is your country?

Country	GNP	Green NNP (\$ per capita 1993)	% fall on GNP
Japan	31,449	27,374	-13.0
Norway	25,947	21,045	-18.9
United States	24,716	21,865	-11.5
Germany	23,494	20,844	-11.3
South Korea	7,681	7,041	-8.3
South Africa	3,582	2,997	-16.3
Brazil	2,936	2,579	-12.2
Indonesia	732	616	-15.8
China	490	411	-16.1
India	293	242	-17.4

Green Net National Product (NNP) is Gross National Product (GNP) minus depreciation of produced assets, depletion of forests and subsoil assets, and damage from carbon dioxide emissions.

zero". But others say a tax encourages polluters to reduce emissions only to the point where the cost of further abatement exceeds the cost of the tax. So the choice between tradable permits and taxes rests on politics and philosophy, rather than economics.

Market-based conservation

Admittedly, many ecosystem services are not yet scarce enough to lend themselves to a market-based conservation mechanism. In other cases, appropriate institutions (or regulations) do not yet exist to assign and enforce the relevant property rights.

But, at least in the United States, where the climate for markets is particularly favourable, there is a wealth of activity in 'enviro-capitalism'. For example, in a trade brokered by the Washington-based Environmental Defense Fund (EDF) in 1993, a large farm in Oregon agreed to lease its rights to divert tens of millions of cubic metres of water from the Snake River to the Bonneville Power Administration (BPA), a federal agency that markets hydropower generated by government-owned dams. The BPA was required by the Endangered Species Act to increase flows at certain times of year to improve conditions for threatened salmon populations. After a three-year pilot project, the trade was made permanent in April 1997, when the Department of the Interior



Pressure point: tradable emissions permits offer one answer to public demands for clean air.

acquired the water rights.

Now the Environmental Resources Trust, a sister organization of the EDF, is marketing the electricity generated by these and other "fish flows" on the Snake and Columbia Rivers as "clean power".

Another straw in the wind comes from elsewhere in the US energy industry. The Electric Power Research Institute (EPRI), a research consortium for the energy industry, is helping one of its member companies,

Allegheny Power, to incorporate the principles of ecological resource management into a land management plan.

Another EPRI member, Southern California Edison, has decided to go into the "conservation banking" business: by agreeing not to develop an ecologically significant site, it earns the right to sell development credits to other parties. So in this industry, at least, economic incentives for conservation are starting to have an effect.

a 'greener' method of calculating national productivity

point, it adds the value of unpaid household work, and then subtracts the cost of pollution, as well as urbanization, road accidents and advertising. But, for all its intellectual attractions, the ISEW has had little success among policymakers. It has also been criticized by more mainstream economists, mainly because of questions about the accuracy of measuring some of its components.

Despite such shortcomings, since the beginning of the decade, calls for a green GDP have been getting louder. In 1993, in response to such suggestions, the United Nations (UN) Statistical Division in New York, which is responsible for setting guidelines for national accounting systems, carried out a review of possible alternatives.

But the review concluded that there was a lack of sufficient data to be able to recommend that countries adopt a green index, or, indeed, a new welfare index.

Instead, countries were encouraged merely to publish separate indicators on the state of key environmental services, as well as their associated monetary values, in parallel with conventional measurements of GDP.

In response to such suggestions, the European Commission is already working on 60 environmental 'pressure indices' intended to act as a measure of the

health of various natural resources.

In parallel, the commission is working with the UN and the Organization for Economic Cooperation and Development (OECD) on so-called 'satellite accounts' that attempt to put monetary values on different aspects of environmental degradation.

The US Environmental Protection Agency has also started a project that seeks to quantify certain ecological services, although this has run into problems in Congress.

There are several other reasons why a green GDP has not been taken up. One is the lack of agreement on its components and on how the index would be calculated.

A second reason is that GDP was never intended to be used as an indicator of environmental health, or indeed of prosperity. And some policymakers see little point in trying to turn it into something that was not originally intended.

Third, there has been unexpected enthusiasm among governments for the Human Development Index (HDI), a quality-of-life indicator that is based on average life expectancy at birth, literacy level, number of years at school and GDP per capita.

Another equally important reason for scepticism about a new green index is the lack of detailed knowledge of its potential components, such as an accurate measure of

water pollution, or the climate change potential of greenhouse gases. The UN review team felt that the work on parallel indicators — which many countries are now carrying out — would help to address most of these issues.

"There was a feeling at the time that if we want to include environmental resources in GDP, we must have comprehensive information," says Kirk Hamilton, a senior economist at the environment department of the World Bank. "That means having detailed knowledge of each type of damage by each pollutant."

Many of these gaps are now being filled. But Hamilton points out that most rich countries remain unconvinced about the desirability of a new index, partly because they derive a smaller share of their earnings from natural resources than developing countries, but also because the parallel indicators are, in themselves, an adequate guide to environmental health.

Despite the setbacks, environmental and ecological economists continue to argue the case for a single index that integrates a measure of the wealth of a country's citizens and the health of its natural environment. Progress towards this goal has been slow. But few have given up hope that it can be achieved in a generally acceptable way.

Audacious bid to value the planet whips up a storm

One of the most controversial recent attempts to integrate economics and ecology has been a calculation, published last year by ecologist Robert Costanza of the University of Maryland and 12 co-authors, of a monetary value for the world's "ecosystem services and natural capital" (see *Nature* 387, 253–260; 1997). Such services include the purification of air and water, the mitigation of floods and drought, pollination, pest control, and the generation of fertile soils.

The idea of valuing ecosystem services was not new. Earlier in the year, the Stanford ecologist Gretchen Daily had published an edited volume, *Nature's Services*, containing contributions that aimed to "identify and characterize components of ecosystem value". But Costanza *et al.* did something that has been described as both "heroic" and "foolhardy": they tried to estimate the total value of all the world's ecosystem services. The answer was \$33 trillion per year: a figure that exceeds the sum of the world's gross national products.

The paper was a box-office success but was panned by the critics. It was widely covered by newspapers and magazines, and the \$33 trillion a year total has been quoted in public speeches by government officials. But many economists characterized the paper as not just wrong but misleading.

Costanza and his colleagues have been resolute in defending the importance of their contribution. And, as the dust settles, it seems that most interested observers believe that a paper with serious technical flaws has still served a useful purpose by drawing attention to an important issue.

Valuing ecosystems

At first glance, what Costanza *et al.* set out to do seems straightforward: namely to rectify the fact that "because ecosystem services are not fully 'captured' in commercial markets or adequately quantified, they are often given too little weight in policy decisions".

The paper describes \$33 trillion per year as "a minimum estimate" for the "current economic value" of 17 ecosystem services (from atmospheric gas regulation to the provision of "cultural value") summed over 16 types of ecosystem, or 'biomes' (from the open ocean to urban centres).

Mainstream economists were quick to protest. In a special issue of *Ecological Economics* devoted to the paper, Michael Toman of Resources for the Future called the \$33 trillion figure "a serious underestimate of infinity". A group of British economists wrote that the biome-scale calculations "risk ridicule from both scientists and economists", and called the figure "not supportable".

Economists complained that Costanza *et al.* didn't properly understand what they were doing. "If you use an economist's definition of valuation, you have to understand what it can be used for and what it can't," says Nancy Bockstael of the University of Maryland.

For neoclassical economists, value can be measured only in the context of a specific exchange. In this view, it is nonsensical to ask "the value" of the world's ecosystem services; an economist would ask, "value to whom?". A related requirement is that one can evaluate only small (or 'marginal') changes from current conditions. Real-world decisions are incremental: we may have to decide what it's worth to give up a hectare of beach, but we are never asked to give up all the beaches in the world.

Local difficulties

Costanza *et al.*, critics complained, strayed out of context by taking valuations of particular ecosystem services, made in specific parts of the world, and converting these to "per hectare" values for a particular biome. For example, they used values placed on soil formation in Colorado as the basis of a calculation for all the world's grasslands. But this assumes that all hectares of grassland are equivalent—not only in their ability to form soil, but in the value of this service to local populations.

The critics also pointed out that, because the value of a commodity increases as it becomes more scarce, one cannot simply multiply the present value of a hectare of biome (even if there were a uniform value per hectare) by the number of hectares to get the total value. The last hectare to disappear will be much more valuable than the first.

In response, Costanza *et al.* argue that what they have done is no different from classical GNP accounting, in which "the total value of marketed products" is computed by multiplying the current price for each product by the number of units traded in a year. But the critics remain unpersuaded.

Costanza treats the detailed criticisms with some impatience, describing himself as a "big picture" person. "This is an order of magnitude study, a first cut," he says. "Probably most economists would have guessed 1 per cent of GNP or less [for the value of ecosystem services]. They're in the wrong order of magnitude. Therefore this issue requires a lot more attention."

One of Costanza's economist co-authors, Stephen Farber, of the Graduate School of Public and International Affairs at the University of Pittsburgh, admits that many of his fellow economists' criticisms are on target. "I don't place a lot of credibility on the \$33 trillion figure," he says. "But if we were

to try to satisfy [our critics in neoclassical economics], doomsday would be past before we got any useful knowledge out there."

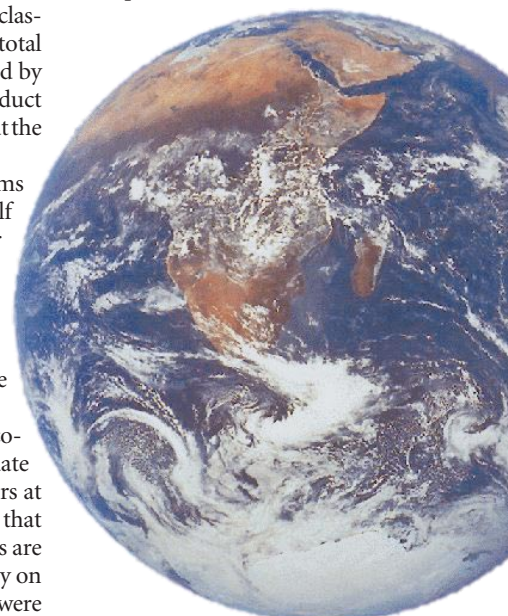
This seems to be the nub of the difference between Costanza *et al.* and their critics. The authors accept that what they did was imperfect in many ways, but feel strongly that their number is better than no number at all.

Ironically, "the number" has come back to haunt one of the economist co-authors, Ralph d'Arge, an emeritus professor from the University of Wyoming, says that he has had "calls from federal agencies asking how they can use this number to implement policy. The answer is that they can't; the per hectare numbers are average—they are unlikely to be a good measure of a local loss." D'Arge nevertheless stands by the paper's methods, saying "we followed all the rules [of neoclassical economics]".

While vocal critics such as Nancy Bockstael view the paper as potentially damaging to their profession, other economists are prepared to take a more philosophical view of its contribution. Trudy Cameron, for example, an environmental economist at the University of California at Los Angeles, characterizes Costanza *et al.*'s paper as "a recklessly heroic attempt to do something that's futile". But then she goes on to say that the paper has been "very useful—it has stirred things up a lot".

Michael Toman, one of the harshest critics of the paper's methods, echoes this view, saying that "it can best be read as a political document". Farber agrees that its main contribution has been in raising awareness of the issues it highlights. "We thought we would provoke, and thought that provocation would be good."

Indeed, the paper's most lasting contribution may be as a recruiting document. "Because of the paper we're seeing young graduate students becoming attached to this issue," says Farber. "Even if people tear the article apart, that's okay if it provokes interest in large-scale problems."



NASA

Plight of US postdocs... in the US

Sir — The Aaron Diamond AIDS Research Center in New York has only a few students, as it is primarily a research institute. But we have many young postdocs. My own lab consists of me (British), six postdocs (one each from Britain, Serbia, China, India and two from Austria), one student (American) and several American technicians.

It is immediately obvious which nationality is missing from the senior line-up. Of the two US postdocs I have had, one left after three years for a position in scientific administration with greater perceived job security, the other quit after a year to move to industry and was paid more for less intense work. I have been trying to recruit another American who has just completed his PhD at a top university. Our institution is his first choice academically, but he prefers to move to industry "because it pays more".

When smart students could go on to make fortunes in US society as lawyers, investors, bankers and so on, how can science compete? All it has going for it is the noble but nebulous "vocational" aspect that

only a minority of students ever feel.

The NIH official stipend for a new postdoc is \$21,000; for students it is a mere \$11,748. In the United States, a PhD in biological sciences is rarely awarded before the age of 30, so smart students spend their 20s earning a pittance compared to their ex-colleagues in other fields. Then they may easily fail to find a full-time academic position, while still having student loans to pay off. Is it any wonder, given the nature of US society, that fewer and fewer young people want to enter science (see *Nature* 395, 101; 1998)? As one of our students put it to me: "when my lawyer friends want to go to a pricey restaurant or I have to spend another sleepless night rushing around the lab, I figure the effort to reward ratio is a bit skewed in science. It's worth it, but sometimes I wonder." The most committed students stick it out regardless, but it is not made easy for them, given alternative temptations.

For foreigners, especially non-Europeans but even including Europeans, US scientific salaries are very attractive by

the standards in their own countries. If it weren't for foreign imports at the postdoc level, the US science structure would rapidly disintegrate. Some countries use the United States to train their brightest young people so that they can create a native scientific infrastructure if and when they finally return: first Japan, then Korea, now China and, to a lesser extent, India. Excluding immigrants is not the answer, unless the United States is willing to minimize its scientific output for several years while it trains home-bred replacements (assuming it could).

In reality, if more young people in the United States are to study science professionally, the government will have to pay them more. Perhaps instead of spending \$40 million on the Starr report, the money could have been used to train more young scientists how to, for example, analyse DNA samples.

John Moore

Aaron Diamond AIDS Research Center,
New York, New York 10016, USA
e-mail: jmoore@adarc.org

Hope or hype for lexipafant?

Sir — I read the News article on the large multicentre trial of lexipafant (Zacutex) in severe pancreatitis and the response of the chief executive of British Biotechnology (*Nature* 393, 291, 299 & 509; 1998). I used lexipafant to block platelet-activating factor in different models of shock and sepsis during 1993–95 at Linköping University, Sweden. It was hard to defend my PhD thesis because of the lack of published data on the basic pharmacology of lexipafant at that time. I am disappointed that this remains the case despite progress in clinical trials.

Commercial considerations resulting from the potential for profit from a new treatment in areas where there is a high rate of death, like severe pancreatitis, are no excuse for keep this basic pharmacological data hidden.

I searched three major medical databases and found that 37 peer-reviewed scientific papers related to lexipafant have been published. Fourteen were review articles and five were original experimental papers on acute pancreatitis, which did not collectively show clear evidence for improved survival (via an anti-inflammatory effect) for those taking this drug. Yet only two of the review articles were conservative about the future of

lexipafant whereas most, including mine, were hopeful.

Scientific literacy in the investment community is of no use without an appreciation of the difference between hope and facts — I will not be surprised if the large multicentre trial of lexipafant in the United States fails, but I hope it will not.

I wonder how much it will cost to acknowledge that the inflammatory process cannot be hit by a single golden bullet. The line between hope and hype is a thin one, and in this case it is only when the results of large-scale trials are available that it will be clear where it should be drawn.

Fikri M. Abu-Zidan

Department of Surgery, Auckland Hospital,
Faculty of Medicine and Health Science,
University of Auckland, Park Road, Auckland,
New Zealand
e-mail: f.abu-zidan@auckland.ac.nz

Animals at the Salk

Sir — I would like to comment on the News item "Salk Institute investigated after claims of inhumane research" (*Nature* 394, 709; 1998). Since this matter is still in litigation, I will not address the specific allegations presented in the article. It is important to note, however, that most of these allegations are part of a continuing civil lawsuit brought by a former Salk employee who was

dismissed for cause. They are not fact and should not be treated as fact.

That said, the first paragraph of your article omits the word "alleged" when describing "past inhumane treatment and faulty experiments" at Salk, while the term "exposed" is also argumentative. Other statements in the article do not accurately represent testimony or court record, or are presented without appropriate caveats. You state that our faculty was divided on the plaintiff's performance, but neglect to inform the reader that only one of 55 supported her and testified on her behalf.

The Salk Institute welcomes an investigation of our animal research department by responsible outside agencies, including the US Department of Agriculture and the NIH. We have run, and continue to run, a first-rate facility that meets all federal guidelines for the humane use of animals in research.

Not only is it in the animals' best interest that we operate in this manner, it is in our own best interests as well. We cannot conduct the kind of research with which we've been entrusted unless our labs and animal facilities are a match for the outstanding scientists we have assembled here.

Thomas D. Pollard

(President)

The Salk Institute for Biological Studies,
Post Office Box 85800
San Diego, California 92186-5800, USA

How economists see the environment

Economists and ecologists misunderstand each other about the environment. Improving interdisciplinary communication should enable natural scientists to take economic analysis and prescriptions more seriously.

Don Fullerton and Robert Stavins

On a topic such as the environment, communication among those from different disciplines in the natural and social sciences is both important and difficult. Economists themselves may have contributed to some misunderstandings about how they think about the environment, perhaps through enthusiasm for market solutions, perhaps by neglecting to make explicit all the necessary qualifications, and perhaps simply by the use of jargon.

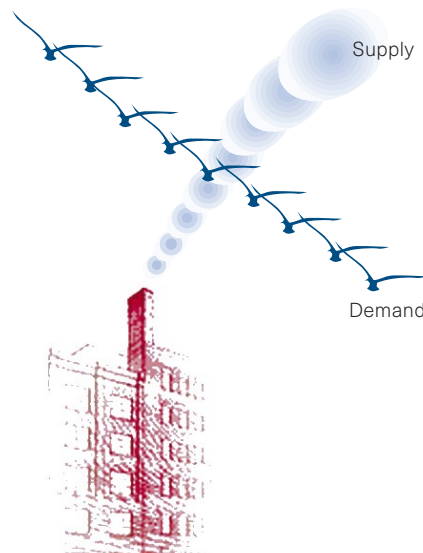
There are several prevalent myths about how economists think about the environment. By examining them here, we hope to explain how economists really do think about the natural environment.

Myth of the universal market

The first myth is that economists believe that the market solves all problems. The "first theorem of welfare economics", as taught to generations of economics students, is that private markets are perfectly efficient on their own, with no interference from government, provided certain conditions are met.

This theorem, easily proved, is exceptionally powerful, because it means that no one needs to tell producers of goods and services what to sell to which consumers. Instead, self-interested producers and consumers meet in the market-place, engage in trade, and thereby achieve the greatest good for the greatest number, as if "guided by an invisible hand"¹. This maximum general welfare is what economists mean by the 'efficiency' of competitive markets. Economists in business schools are particularly fond of identifying markets where the necessary conditions are met, such as the stock market, where many buyers and sellers operate with good information and low transaction costs to trade well-defined commodities with enforced rights of ownership.

Other economists, especially those in public policy schools, have a different approach to this theorem. By clarifying the conditions under which markets are efficient, the theorem also identifies the conditions under which they are not. Private markets are perfectly efficient only if there are no public goods, no externalities, no monopoly buyers or sellers, no increasing returns to scale, no information problems, no transaction costs, no taxes, no common property and no other 'distortions' between the costs paid by buyers and the benefits received by sellers. Those conditions are obviously very restrictive, and they are usually not all satis-



An economist's view of the environment.

fied simultaneously in the real world.

When a market thus fails, this same theorem offers guidance. For any particular market, it asks whether the number of sellers is sufficiently small to warrant antitrust action, whether the returns to scale are great enough to justify tolerating a single producer in a regulated market, or whether the benefits from the good are public in a way that might justify outright government provision of it. A public good, like the light from a lighthouse, benefits additional users at no cost to society.

Environmental economists are interested in pollution and other externalities, where some consequences of producing or consuming a good or service are external to the market (not considered by producers or consumers). With a negative externality, such as environmental pollution, the total social cost of production may exceed the value to consumers. If the market is left to itself, too many pollution-generating products are made.

Similarly, natural-resource economists are interested in common property, or open-access resources, where anyone can extract or harvest the resource freely and no one recognizes the full cost of using the resource. Extractors consider only their own direct and immediate costs, not the costs to others of increased scarcity ('user cost' or 'scarcity rent'). The result is that the resource is depleted too quickly.

So, the market by itself demonstrably does not solve all problems. Indeed, in the environmental domain, perfectly functioning markets are the exception rather than the rule. Governments can try to correct these market failures, for example by restricting

pollutant emissions or limiting access to open-access resources, which can improve welfare and lead to greater efficiency.

Myth of market solutions

A second common myth is that economists always recommend a market solution to a market problem. Economists tend to search for instruments of public policy that can fix one market essentially by introducing another, allowing each to operate efficiently on its own. If pollution imposes large external costs, for example, the government can establish a market for rights to emit a limited amount of that pollutant. Such a market for tradable emission permits will work if there are many buyers and sellers, all are well informed, and the other conditions of the 'first theorem' are met. In this case, the government's role is to enforce the rights and responsibilities of permit ownership, so that each unit of emissions is matched by the ownership of one emission permit. Then the market for the output will also work, as the producer has to pay a price for each permit that reflects the social cost of the associated pollution. Equivalently, producers can be required to pay a tax on their emissions that reflects the external social cost. Either way, the result in theory will be the efficient amount of pollution abatement, undertaken at minimum aggregate abatement cost.

This tradable-permit approach has much to recommend it, and can be just the right solution in some cases, but it is still a 'market'. Therefore the outcome will be efficient only if certain conditions are met. But these conditions are not always met². Could the sale of permits be monopolized by a small number of buyers or sellers? Do problems arise from inadequate information or significant transaction costs? Will the government find it too costly to measure emissions? If the answer to any such question is yes, the permit market may work less than optimally. The environmental goal may still be met, but at more than minimum cost.

As an example, to reduce acid rain in the United States, amendments to the Clean Air Act of 1990 require electricity generators to hold a permit for each tonne of SO₂ they emit. A robust market for the permits has emerged, in which well-defined prices are broadly known to many potential buyers and sellers. Through continuous emissions monitoring, the government can track SO₂ emissions from each plant. Equally important, penalties are significantly greater than incremental abatement costs and hence are sufficient to ensure compliance. Overall, this market

works; acid rain deposition is being reduced by 50 per cent in a cost-effective manner.

A permit market achieves this efficiency through trades because any company that has high abatement costs can buy permits from another that has low costs, so reducing the total cost of abating pollution. These trades also switch the source of the pollution from one company to another, which is unimportant when any emissions equally affect the whole trading area. This 'perfect mixing' assumption is certainly valid for global problems such as greenhouse gases or the effect of chlorofluorocarbons on the stratospheric ozone layer. It may also work reasonably well for a regional problem such as acid rain, because acid deposition in downwind states of New England is about equally affected by SO₂ emissions that were traded among upwind sources in Ohio, Indiana or Illinois. But it does not work perfectly, as acid rain in New England may increase if a plant there sells permits to a plant in the mid-west.

At the other extreme, many environmental problems might not be addressed appropriately by tradable-permit systems or other market-based policy instruments⁴. One example is a hazardous air pollutant such as benzene that does not mix in the airshed and so can cause localized 'hotspots'. Because a company can buy permits and increase local emissions, permit trading does not ensure that each location will meet a specific standard. Moreover, the damages caused by local concentrations may increase nonlinearly. If so, then even a permit system that reduces total emissions might allow trades that move those emissions to a high-impact location and thus increase total damages.

The bottom line is that no specific policy instrument, or even set of policy instruments, is a panacea. Market instruments do not always provide the best solutions, and sometimes not even satisfactory solutions.

Myth of market prices

The next myth is that, when non-market solutions are considered, economists still use only market prices to evaluate them. No matter what policy instrument is chosen, the environmental goal of that policy must be identified. For example, should vehicle emissions be reduced by 10, 20 or 50 per cent? Economists frequently try to identify the most efficient degree of control that provides the greatest net benefit. This means, of course, that both benefits and costs need to be evaluated. True enough, economists typically favour using market prices, whenever possible, to carry out such evaluations, because these prices reveal how members of society actually value the scarce amenities and resources under consideration.

Economists are wary of asking people how much they value something, as respondents may not provide honest assessments of their own valuations. Instead, actions may

reveal their preferences, as when individuals pay more for a house in a neighbourhood with cleaner air, all else being equal⁵.

This is not to suggest that economists are concerned only with the financial value of things. Far from it. The financial flows that make up the gross national product represent only a fraction of all economic flows. The scope of economics encompasses the allocation and use of all scarce resources. For example, the economic value of the human-health damages of environmental pollution is greater than the sum of health-care costs and lost wages (or lost productivity), as it includes what lawyers would call 'pain and suffering'. Economists might use a market price indirectly to measure revealed rather than stated preferences, but the goal is to measure the total value of the loss that individuals incur.

To take another example, the economic value of part of the Amazon rainforest is not limited to its financial value as a repository of future pharmaceutical products or as a location for ecotourism. That 'use' value may only be a small part of the properly defined economic valuation. For decades, economists have recognized the importance of 'non-use' value of environmental amenities such as wilderness areas or endangered species. The public nature of these goods make it particularly difficult to quantify these values empirically, as we cannot use market prices! The important fact is that benefit-cost analysis of environmental policies, virtually by definition, cannot rely exclusively on market prices⁶.

Economists insist on trying to convert all these disparate values into monetary terms because a common unit of measure is needed to be able to add them up. How else can we combine the benefits of ten extra miles of visibility plus some amount of reduced morbidity, and then compare these total benefits with the total cost of installing scrubbers to clean stack gases at coal-fired power plants? Money, after all, is simply a medium of exchange, a convenient way to add together or compare disparate goods and services.

Myth of efficiency

The last myth we address here is that these economic analyses are concerned only with efficiency rather than distribution. Many economists do give more attention to measures of aggregate social welfare than to measures of the distribution of the benefits and costs of policies among members of society. The reason is that an improvement in economic efficiency can be determined by a simple and unambiguous criterion — an increase in total net benefits. What constitutes an improvement in distributional equity, on the other hand, is inevitably the subject of considerable dispute. Nevertheless, many economists do analyse distributional issues thoroughly. The more difficult problem, not yet solved in a satisfactory manner, is how to

combine efficiency and distributional issues in a unified analysis.

Available data often permit reliable estimates of the impacts of environmental policies on important subgroups of the population⁷. On the other hand, environmental regulations are neither effective nor efficient tools for achieving redistributional goals. The best economic analyses recognize the contributions and limitations of efficiency and distributional measures.

Where does this leave us?

To summarize, economists do not necessarily believe that the market solves all problems. Indeed, many economists, ourselves included, make a living out of analysing market failures such as environmental pollution in which laissez-faire policy leads not to social efficiency, but to inefficiency. When economists identify market problems, their tendency is first to consider the feasibility of market solutions because of their potential cost-effectiveness, but market-based approaches to environmental protection are no panacea. When market or non-market solutions to environmental problems are being assessed, economists do not limit their analysis to financial considerations but use money as a unit of measurement in the absence of a more convenient unit. And although the efficiency criterion is by definition aggregate in nature, economic analysis can reveal much about the distribution of the benefits and costs of environmental policy.

Having identified and sought to dispel four prevalent myths about how economists think about the natural environment, we acknowledge that our profession bears some responsibility for the existence of such misunderstandings. Like their colleagues in other social and natural sciences, academic economists focus their greatest energies on communicating to their peers within their own discipline. Greater effort can certainly be made to improve communication across disciplinary boundaries. □

Don Fullerton is in the Department of Economics, University of Texas at Austin, Texas 78712, USA and at the National Bureau of Economic Research. Robert Stavins is at the John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA and Resources for the Future.

The authors are grateful for suggestions from Robert Frosch, Robert Hahn, Gilbert Metcalf, Richard Revesz and Thomas Schelling.

1. Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations* (Whitstone, Dublin, 1776).
2. Hahn, R. W. & Hester, G. L. *Ecol. Law Q.* **16**, 361–406 (1989).
3. Schmalensee, R. *et al. J. Econ. Perspect.* **12**, No. 3 (Summer 1998).
4. Hahn, R. W. & Stavins, R. N. *Am. Econ. Rev.* **82**, 464–468 (1992).
5. Smith, V. K. & Huang, J.-C. *J. Polit. Econ.* **103**, 209–227 (1995).
6. Arrow, K. *et al. Science* **272**, 221–222 (1996).
7. Christiansen, G. B. & Tietenberg, T. H. in *Handbook of Natural Resource and Energy Economics* Vol. 1 (eds Kneese, A. V. & Sweeney, J. L.) 345–393 (North-Holland, Amsterdam, 1985).

Kepler's conjecture confirmed

Neil J. A. Sloane

One of the oldest unsolved problems in mathematics appears to have been settled. On 9 August, Thomas C. Hales announced that he had proved Kepler's assertion of 1611 that no packing of spheres can be denser than a face-centred-cubic lattice.

In face-centred-cubic packing (Fig. 1), seen in the piles of oranges in any grocer's shop, the spheres occupy 0.7405 of the total space available. Ambrose Rogers remarked¹ in 1958 that "many mathematicians believe and every physicist knows" that no denser packing is possible. So why has it taken 387 years for a proof to be found?

There are three reasons. First, technical difficulties come from the fact that the density of a packing is defined as the *limit* of the fraction of space occupied by the balls as the number of balls goes to infinity. This means that (say) a million balls can be thrown away without changing the density.

Second, even if one considers only packings without any obvious gaps, there are still infinitely many different packings that are just as dense as the face-centred-cubic lattice. These are the 'Barlow packings', described in this journal² in 1883: put down a layer of spheres arranged in the triangular lattice (the arrangement used when racking billiard balls), place another layer on top, and repeat. There are two ways to place each layer after the second (Fig. 2), giving an uncountable infinity of distinct packing schemes, all with the same density. Third, in the face-centred-cubic lattice, each ball touches 12 others. But there are infinitely many distinct ways to arrange 12 balls around another ball of the same size, because there's a lot of slack — almost room to squeeze in a thirteenth ball.

It is these various kinds of non-uniqueness that make the problem hard. (Paradoxically, the sphere-packing problems in four and eight dimensions may turn out to be easier, because the last two types of non-uniqueness are absent.) So far, no one has found a simple approach to the three-dimensional problem.

The methods that have been tried, and which in Hales's hands finally succeeded, are messy. The basic approach is to reduce a minimization problem involving infinitely many variables to a finite number of sub-problems, and until the proof is finished it is never certain that the splitting into sub-problems will terminate.

There have been two unsuccessful attacks on the problem in recent years that received widespread attention. Buckminster Fuller claimed to have a proof³ in 1975, but his



Figure 1 Cannonballs stacked in a face-centred-cubic lattice (Arlington, Virginia, about 1863). There is no denser way to do it.

arguments amount to a description of the face-centred-cubic packing rather than a proof of its optimality. Then in 1993 Wu-Yi Hsiang published a 100-page paper⁴ claiming to give a proof. But there are serious flaws in his arguments too. To quote one review⁵: "If I am asked whether the paper fulfils what it promises in its title, namely a proof of Kepler's conjecture, my answer is: no. I hope that Hsiang will fill in the details, but I feel that the greater part of the work has yet to be done."

Hales has been working on the Kepler conjecture for ten years, having proposed a five-step attack on the problem. Two parts have already been refereed and published⁶. The now completed proof⁷ relies heavily on computers, which are used in several different ways. For example, at the heart of the proof are some 100,000 linear programming problems, each involving 100 to 200 variables and 1,000 to 2,000 constraints. Hales has been careful to make all the computer calculations available on his website⁷. Although the final acceptance of his claim must await a careful refereeing of the proof, there is little reason to doubt it. There was never any strong reason to suspect that Kepler's conjecture was false — although it was a little worrisome that there are packings⁸ of identical ellipsoids that have densities exceeding 0.75 — but the proof is a considerable achievement.

Far from being the last word, Hales's result is just the beginning. Communication

theorists as well as mathematicians are interested in determining the densest sphere packings in dimensions above three. The sampling theorem of information theory⁹ says that a signal containing no frequencies above W hertz can be reconstructed from samples taken every $1/(2W)$ seconds. So a signal that lasts for T seconds can be represented by $2WT$ samples. Just as three numbers specify the coordinates of a point in three-dimensional space, so these $2WT$ samples specify a point in $2WT$ -dimensional space. The whole waveform is specified by a single point in $2WT$ -dimensional space. Similar signals are represented by nearby points, dissimilar signals by well-separated points. So one of the fundamental problems in communication theory is determining the densest packing of balls in high-dimensional spaces.

This geometrical way of representing signals, at the heart of Shannon's mathematical theory of communication⁹, underlies the high-speed modems that we now take for granted. One of the most common coding schemes in use today works so well because the signals are represented as points in eight-dimensional space.

Many beautiful packings are known in high dimensions, and have fascinating and unexpected connections with other branches of mathematics¹⁰. John H. Conway and I have described¹¹ what we think may be all the best sphere packings in dimensions up to ten, where 'best' means that they have the highest density and contain no obvious flaws such as cavities or cracks.

Let me conclude by mentioning an amazing property of what we conjecture are the densest nine-dimensional packings. In some of these schemes (again there are infinitely many that are equally dense), half the spheres can be moved bodily through arbitrarily large distances without overlapping the other half, only touching them at isolated instants — all without changing

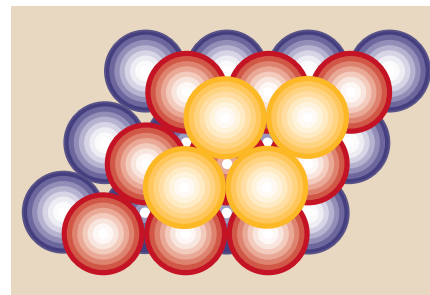


Figure 2 Stacking schemes. Rack up a layer of billiard balls (purple); place another layer on top of them (red). The third layer can fall in a third position (yellow) as shown, or alternatively can sit above the purple spheres. You also have two choices for every subsequent layer, so there are infinitely many packings as dense as the face-centred-cubic pattern² (which corresponds to repeated purple, red, yellow layers).

the density of the packing.

But until someone extends Hales's result to higher dimensions, we have no proof that any packing in a dimension greater than three is optimal. □

Neil J. A. Sloane is at the AT&T Shannon Laboratory, 180 Park Avenue, Florham Park, New Jersey 07932-0971, USA.
e-mail: njas@research.att.com

1. Rogers, C. A. *Proc. Lond. Math. Soc.* **8**, 609–620 (1958).
2. Barlow, W. *Nature* **29**, 186–188 (1883).

3. Buckminster Fuller, R. *Synergetics* (Macmillan, London, 1975).
4. Hsiang, W.-Y. *Int. J. Math.* **93**, 739–831 (1993).
5. Fejes Tóth, G. *Math. Rev.* **95**, no. 52032 (1995).
6. Hales, T. C. *Discrete Computational Geom.* **17**, 1–51 (1997); **18**, 135–149 (1997).
7. <http://www.math.lsa.umich.edu/~hales/countdown>
8. Bezdek, A. & Kuperberg, W. in *Victor Klee Festschrift* (eds Gritzmann, P. & Sturmfels, B.) 71–80 (Am. Math. Soc., Providence, RI, 1991).
9. Shannon, C. E. *Proc. Inst. Radio Engineers* **37**, 10–21 (1949).
10. Conway, J. H. & Sloane, N. J. A. *Sphere Packings, Lattices and Groups* 2nd edn (Springer, New York, 1993).
11. Conway, J. H. & Sloane, N. J. A. *Discrete Computational Geom.* **13**, 383–403 (1995).

sponding specialized types of segment⁴.

Damen *et al.* and Telford and Thomas report the identification and expression of Hox genes in two different members of the Chelicerata, a spider¹ and a mite². The results from the two species are congruent and point to the same conclusion — that expression of Hox genes in the prosoma of chelicerates is similar to the expression of homologous genes in the head segments of other arthropods. This suggests that the entire prosoma of chelicerates may correspond to a head (Fig. 2). Indeed, the similarities are so striking that the authors do not hesitate to suggest specific one-to-one homologies between segments of the prosoma and the head segments of other arthropods.

Most convincing are those of the most anterior parts of the body, between the cheliceral segment and the first antennal (deutocerebral) segment, and between the pedipalpal and the second antennal or intercalary (tritocerebral) segment. In more

Evolutionary biology

Origin of the spider's head

Michalis Averof

Comparing the structure of the head between different classes of arthropod has been controversial and often frustrating, yet it can tell us about the origins and relatedness of these different classes. The heads of crustaceans, myriapods and insects contain a characteristic set of at least five segments, which are thought to be homologous among these groups. But the morphology of chelicerates — which include spiders, scorpions, mites and horseshoe crabs — gives no good clues about the relationship of their segments to the head segments of other arthropods. Now, however, studies based on the expression of developmental genes by Damen *et al.*¹ and Telford and Thomas², published in *Proceedings of the National Academy of Sciences*, provide convincing evidence for previously unexpected homologies.

Chelicerates do not have a body region that could be obviously characterized as a head. Their bodies are subdivided into two portions — the prosoma (front) and the opisthosoma (back). The prosoma contains segments that bear the chelicerae (spider's fangs), pedipalps and four pairs of walking legs. It carries the main feeding, sensory and locomotor apparatus and, in a functional sense, acts as a head and thorax combined (Fig. 1). But from its morphology, we have so far been unable to find clear evidence for the relationship between this region and the segments of other arthropods.

Damen *et al.*¹ and Telford and Thomas² now provide evidence that comes not from the morphology itself, but from the genes that generate that morphology. They studied the Hox (homeotic) genes, which are known to specify the identity of the different types of segment within the body of insects³. These genes are particularly attractive for distant evolutionary comparisons of segment specialization^{3–5} because their function seems to be widely conserved (in animals as diverse as arthropods, vertebrates and nematodes), and their activity is faithfully reflected by restricted expression in particular regions of the body. Different classes of arthropod have almost identical sets of Hox genes^{3,5}, but

these are expressed differently in animals that have distinct patterns of segmental specialization^{4–6}. Similarities in the expression of these genes have been used as evidence for a common origin (homology) of the corre-



Figure 1 The prosoma of a horseshoe crab (ventral view). The mouth is surrounded by legs bearing prominent endites (arrowheads), a characteristic associated with feeding. Damen *et al.*¹ and Telford and Thomas² propose that the prosoma corresponds, in evolutionary terms, to the head of other arthropods such as crustaceans and insects.

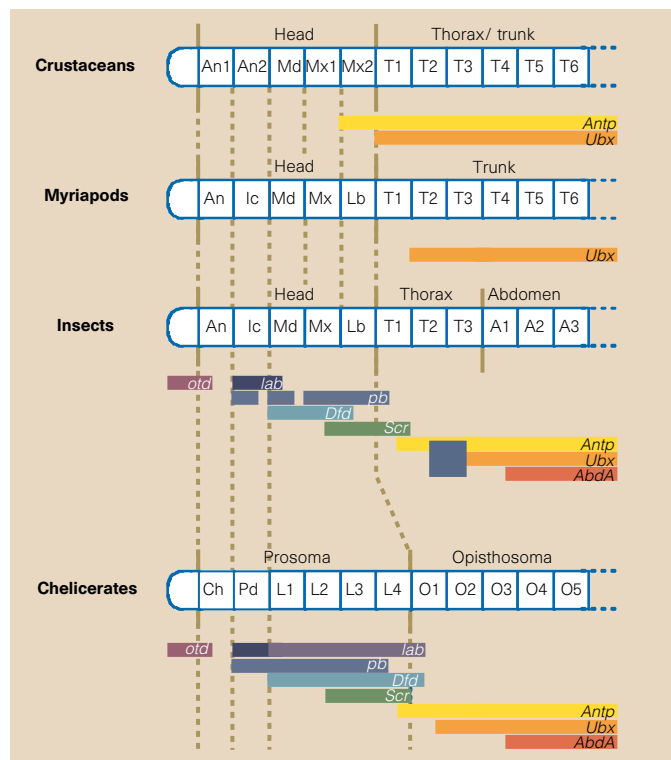


Figure 2 Expression of Hox genes and inferred relationships among anterior body segments in different arthropods. Arthropods express almost identical sets of Hox genes in different regional domains (colour coded for homologous genes)^{1–6,9}. Although there are some variations within arthropod classes, the expression domains shown are those thought to be typical or ancestral for each class. The homologies between insect, crustacean and myriapod segments are based on morphological evidence.

posterior segments, the assignment of homologies becomes more doubtful, as the patterns of morphological diversification (for example, the distinction between prosomal and opisthosomal segments in chelicerates versus head and trunk/thoracic segments in other arthropods) do not correspond so precisely between different arthropods. Nevertheless, it seems likely that the leg-bearing region of the prosoma is related to the posterior head (gnathal) region of crustaceans, myriapods and insects, and that the limbless opisthosomal region might correspond to the thorax or trunk of the other arthropods.

Apart from these similarities, some differences in the expression of Hox genes can be found in the transitional region around the posterior part of the head/prosoma and the anterior part of the trunk/opisthosoma. The significance of these differences is not yet clear. Shifts in the expression of Hox genes have been associated with evolutionary changes in the morphology and function of particular types of segment⁶, but this is not easy to assess in animals with widely divergent morphology. Cambrian fossils show that there were different patterns of segmental specialization in that region of the body^{7,8}, suggesting that segmental identities in that region have not stayed unchanged.

Comparing patterns of gene expression provides a new way to study morphology. Although there is no reason to assume that data from gene expression will be superior to

anatomical data — they both reflect different aspects of the same reality, the evolution of genetically encoded morphological features — these comparisons are attracting much interest today. The hope is that this so-far-unexploited source of information might help us to resolve some of the tricky or unexpected relationships that remain concealed by the external appearance. Chelicerates and the other arthropods are separated by about 550 million years of evolution. During that time, crustaceans, myriapods and insects have incorporated a number of anterior segments into their head and modified some of the corresponding appendages to function as jaws. The new data indicate that spiders make surprisingly different use of these appendages — they use them for walking. □
Michalis Averof is at the Institute of Molecular Biology and Biotechnology, FORTH, Vassilika Vouton, 711 10 Iraklio, Crete, Greece.
e-mail: averof@imbb.forth.gr

1. Damen, W. G. M., Hausdorf, M., Seyfarth, E. A. & Tautz, D. *Proc. Natl Acad. Sci. USA* **95**, 10665–10670 (1998).
2. Telford, M. J. & Thomas, R. H. *Proc. Natl Acad. Sci. USA* **95**, 10671–10675 (1998).
3. Akam, M. *Phil. Trans. R. Soc. Lond. B* **349**, 313–319 (1995).
4. Averof, M. & Akam, M. *Nature* **376**, 420–423 (1995).
5. Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M. & Carroll, S. *Curr. Biol.* **7**, 547–553 (1997).
6. Averof, M. & Patel, N. H. *Nature* **388**, 682–686 (1997).
7. Delle Cave, L. & Simonetta, A. M. in *Early Evolution of Metazoa and the Significance of Problematic Taxa* (eds Simonetta, A. M. & Conway Morris, S.) 189–244 (Cambridge Univ. Press, 1991).
8. Walossek, D. & Müller, K. J. in *Arthropod Relationships, Systematics Assoc. Spec. Vol. Ser. 55* (eds Fortey, R. A. & Thomas, R. H.) 139–153 (Chapman & Hall, London, 1997).
9. Mahaffey, J. W., Diederich, R. J. & Kaufman, T. C. *Development* **105**, 167–174 (1989).

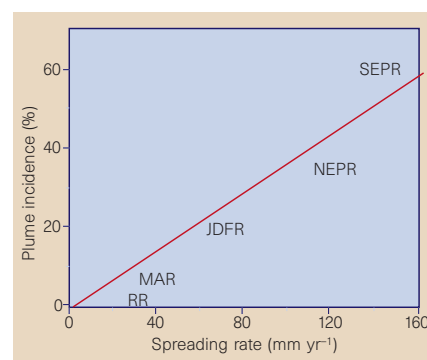


Figure 1 Incidence of hydrothermal plumes plotted against spreading rate of ocean ridges, where plume incidence is the percentage of ridge-axis length that is overlain by significant plume anomalies in the water column. The red square is the predicted value for extremely slow-spreading ridges (<15 mm yr⁻¹) such as the southwest Indian ridge. This model, based on spreading rate alone, underestimates the frequency of hydrothermal venting on the southwest Indian ridge by three times, as German *et al.*¹ now show. RR, Reykjanes ridge; MAR, Mid-Atlantic Ridge; JDFR, Juan de Fuca ridge; NEPR, northern East Pacific Rise; SEPR, southern East Pacific Rise. (Modified from ref. 4.)

plumes, and concluded that plume incidence increases linearly with spreading rate (Fig. 1). This model predicts that the ultra-slow end-member exemplified by the southwest Indian ridge (14 mm yr⁻¹) should have no more than one vent site every 200–300 km. German *et al.*¹ have now tested this prediction with extensive plume surveys and find it to be off by more than a factor of three.

In explaining the higher-than-expected incidence of hydrothermal vents on extremely slow-spreading ridges, German *et al.* implicate tectonism as the process that allows sea water to penetrate to deep heat sources and sustain hydrothermal activity at spatial frequencies that are higher than those predicted from spreading rate alone. To an observer in a submersible, hydrothermal activity at fast-spreading centres is patently influenced by volcanism: basalt lavas at the spreading axis are barely fractured and fissured by tectonics before the next incidence of combined tectonism and volcanism refreshes the surface and incrementally spreads the plates apart. On slow-spreading centres, it is the deep ravages of tectonism that are most obvious, while spreading increments involving volcanic eruptions are infrequent.

The discovery of numerous vents on the southwest Indian ridge opens the door for site visits by geologists who wish to understand controls on hydrothermal activity, and by geochemists intent on parsing and integrating estimates of thermal and chemical fluxes through hydrothermal systems. Biologists, too, will jump at a chance to visit these vents, with the prospect of discovering new

Oceanography

Vents at higher frequency

Cindy Lee Van Dover

The deep-ocean research community is a small one, but the rate at which central dogmas are toppled must be one of the highest in science. Because of the relative inaccessibility of the abyssal environment, general ideas about deep-ocean processes are often advanced on the basis of only a few local or regional observations. As access to the sea floor increases with developing technologies and oceanographic infrastructures, the science becomes ever more volatile and exciting. An example of work that causes us to think again comes in the paper by German *et al.* on page 490 of this issue¹ — there the authors provide remote-sensing data that challenge the current model correlating frequency of hydrothermal vents with the spreading rate of mid-ocean ridges.

When hydrothermal vents were first discovered along submarine mountain ranges in the eastern Pacific, geologists pointed to the obvious link between these hot springs and volcanism. The ensuing searches for vents concentrated on thermally swollen crust of local topographic highs along the

mid-ocean ridge axes. The first vents were found on fast-spreading centres — the East Pacific Rise and Galapagos rift — where the ocean plates move apart at spreading rates of 50–80 mm yr⁻¹. At the time, many scientists expressed doubt that hydrothermal activity could be supported on the Mid-Atlantic Ridge, where spreading is much slower (20–40 mm yr⁻¹) and the thermal and volcanic budget is considerably lower. Peter Rona and his colleagues proved this idea wrong in the mid-1980s with the discovery of the so-called TAG site², which still ranks as one of the largest and most persistent hydrothermal sites yet discovered on the sea floor.

Ten years later, Baker *et al.*³ reported the results of sampling hundreds of kilometres of spreading centres moving at various rates — slow, intermediate, fast and ultrafast (>80 mm yr⁻¹). They used vent-prospecting techniques based on two- and three-dimensional mapping of the thermal, optical and chemical anomalies in the water column that are associated with near-bottom, black smoker

species, physiologies and ecologies, and adding further pieces to the puzzles of global biogeography and evolution of vent faunas. We know that vent faunas are regionally disparate; each sampling of remote sites yields dozens to hundreds of new species, and insight into their distribution and evolution.

Equally exciting is the fact that German and colleagues' findings mean that nowhere is sacred. We can expect to find hydrothermal vents in extremely isolated ridge segments, despite their slow spreading rates. High on the wish-list of vent biogeographers is access to a vent site on the Cayman rise in the Caribbean, where the spreading rate is even slower than on the southwest Indian ridge. Will the fauna of a Cayman vent be Atlantic or Pacific in nature? Or will it be more closely related to seep faunas of the Gulf of Mexico? What of the fauna of slow-spreading Arctic ridge hydrothermal vents, which are located in a basin where the deep water has long been isolated from the major oceans? And what of the putative vent fauna of the Scotia ridge, that island of a spreading

centre that lies well west of the southern Mid-Atlantic Ridge?

It is to these remote ridge systems that I look for vindication of what I think of as the 'trilobite factor' — the belief that there are still major discoveries waiting for us in unexplored regions of the ocean. Trilobites, of course, have been extinct since the Permian, some 250 million years ago, and the chances of finding some deep-water relative of these fossil arthropods teeming about an abyssal hot spring are perhaps remote to nil. But, if not trilobites, surely other unimagined celebrations of nature's ingenuity will be found. □

Cindy Lee Van Dover is in the Department of Biology, Millington Hall, College of William and Mary, Williamsburg, Virginia 23185, USA.
e-mail: cindy_vandover@wm.edu

1. German, C. R., Baker, E. T., Mevel, C., Tamaki, K. & the FUJI Science Team *Nature* **395**, 490–493 (1998).
2. Rona, P. A., Klinkhammer, G., Nelsen, T. A., Trefry, J. H. & Elderfield, H. *Nature* **321**, 33–37 (1986).
3. Baker, E. T., Chen, Y. J. & Morgan, J. P. *Earth Planet. Sci. Lett.* **142**, 137–145 (1996).
4. Baker, E. T., German, C. R. & Elderfield, H. in *Seafloor Hydrothermal Systems: Physical, Chemical, Biological and Geological Interactions* (eds Humphris, S. et al.) 47–71 (Am. Geophys. Un., Washington DC, 1995).

Protein transport

The doors to organelles

Gottfried Schatz

Mitochondria import most of their proteins from the surrounding cytosol. During this process, proteins destined for import are first selected from the thousands of cytosolic proteins and are then transported into or across the outer membrane of the organelle. Finally, many of them must be sorted to their correct destination inside. Reports by Hill *et al.*¹ (page 516 of this issue) and Künkele *et al.*² (in *Cell*) now provide exciting information on the molecular architecture of the channels that transport proteins through the outer membranes of mitochondria.

The protein-transport channel in the yeast mitochondrial outer membrane was first identified by crosslinking one of its subunits to an artificial precursor protein stuck in the channel³. The subunit, now termed Tom40 (ref. 4), is an integral membrane protein (relative molecular mass, M_r , 40,000) that is essential for viability of yeast cells. Tom40 is part of a large 'Tom complex' (Fig. 1) that also contains the small, membrane-embedded subunits Tom5, Tom6 and Tom7, as well as subunits with cytosolically exposed domains (Tom20, Tom22, Tom37 and Tom70). Tom40 is buried in the mitochondrial outer membrane, which it seems to span almost exclusively in the form of β -strands⁵. This topology resembles that of the pores in bacterial outer membranes, so Tom40 was suspected to form the core of the protein-transport channel.

Hill *et al.*¹ now provide direct and convincing evidence for this hypothesis. They overproduced yeast mitochondrial Tom40 in *Escherichia coli* and purified it by denaturation followed by renaturation. They then incorporated the pure Tom40 asymmetrically into liposomes or planar phospholipid bilayers, and showed that the reconstituted protein forms a hydrophilic, cation-selective transmembrane channel. The channel is partly blocked by a peptide representing a functional mitochondrial targeting sequence, but only if the peptide is added to the side of the artificial membrane that corresponds to the cytosolic face of the mito-

chondrial outer membrane. Calculations based on the channel conductance indicate a pore diameter of 12 to 26 Å, depending on the assumed resistivity on the channel — when the authors calibrated the channel with differently sized cations they found a pore diameter of 22 Å. This value agrees with observations that the channel is big enough to allow the passage of precursors that are attached to double-stranded DNA⁶. Furthermore, measurements of the Tom channel with calibrated gold particles yield a value of greater than 22 Å but less than 26 Å (A. Matouschek, personal communication).

What does the protein-import machinery look like in the mitochondrial outer membrane? In a remarkable *tour de force*, Künkele *et al.*² have purified the whole Tom complex from mitochondria of the fungus *Neurospora crassa*, determined its subunit stoichiometry, and verified its functional integrity by reconstitution experiments similar to those reported by Hill and colleagues. Künkele *et al.* also examined the morphology of the channel by negative staining and high-resolution electron microscopy. The filtered images reveal particles with one, two or three stain-filled holes. The 20-Å diameter of the holes agrees reasonably well with the ion conductance of the reconstituted complex if some additional assumptions are made. If each stain-filled hole were indeed a protein channel, an intact Tom complex would have three protein channels.

Whereas subunits of the protein-transport system in mitochondria are called Tom, in chloroplasts they are known as Toc (ref. 7), although no Toc subunit shows any sequence similarity to any Tom subunit. Proteins enter chloroplasts through the Toc complex in the outer membrane (Fig. 1). Toc86 (which is now thought to be a proteolytic fragment of a protein, M_r 159,000, and should be renamed Toc159; D. Schell, personal communication) and Toc34 seem to be GTP-regulated import receptors; the function of Toc36 is still under debate; and Toc75 is thought to form the protein-

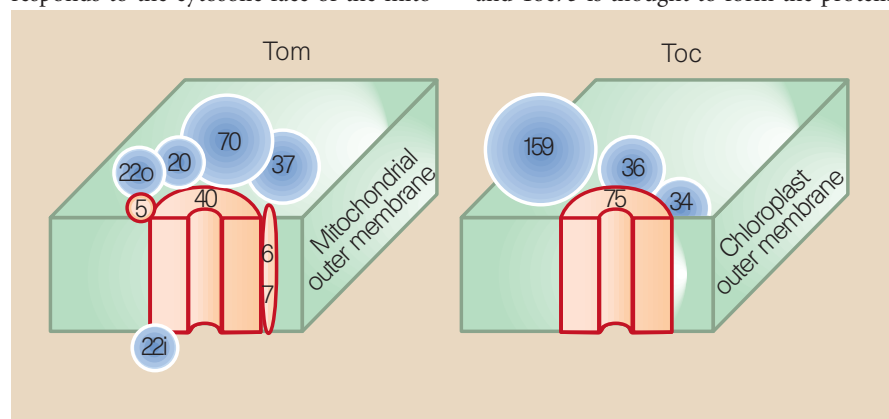


Figure 1 The protein-transport channels in the outer membrane of mitochondria (Tom) and chloroplasts (Toc). Tom22 has extra-membrane domains (Tom22o and Tom22i) on both sides of the outer membrane. The role of Toc36 is still uncertain. The drawing does not indicate the number of each subunit per complex.

transport channel because it is almost completely embedded in the membrane, and is predicted to span the outer membrane through 16 β -strands⁸. In a seminal study, Hinnah *et al.*⁹ incorporated recombinantly expressed and purified Toc75 into liposomes or planar phospholipid bilayers and found that the protein formed voltage-dependent ion channels with weak cation selectivity. The channel was partly blocked by a chloroplast precursor protein, but not by the corresponding mature protein. As in Hill and colleagues' study, the precursor protein inhibited the channel only if it was added to the side of the membrane corresponding to the cytosolic face of the chloroplast outer membrane. The conductance of the Toc75 channel suggested a pore diameter of only 8 Å, but this was not corrected for possible channel resistivity and is probably a minimal value. So despite their different primary sequences, Toc75 and Tom40 have a similar transmembrane topology and function.

The protein-transport channels across another organelle, the endoplasmic reticulum, open only in response to an appropriate precursor protein — they are gated¹⁰. Gating must also be inferred for the corresponding channels across the proton-impermeable inner membranes of mitochondria, chloroplasts and Gram-negative bacteria.

Although there is no direct evidence that protein channels in the outer membranes of mitochondria and chloroplasts are gated, Hill *et al.*¹ find that the Tom40 channel has a three- to fourfold lower probability of opening if it is reconstituted from total outer membranes instead of from purified Tom40. Gating may be effected by the small Tom subunits, which have already been shown to modulate the flexibility of the Tom channel¹¹. In sum, we know the door through which proteins enter mitochondria and chloroplasts, but must still find the parts that make it a gate. □

Gottfried Schatz is in the Department of Biochemistry, Biozentrum, Basel University, Klingelbergstrasse 70, CH-4056 Basel, Switzerland. e-mail: schatz@ubaclu.unibas.ch

1. Hill, K. *et al.* *Nature* **395**, 516–521 (1998).
2. Künkele, K.-P. *et al.* *Cell* **93**, 1009–1019 (1998).
3. Vestweber, D., Brunner, J., Baker, A. & Schatz, G. *Nature* **341**, 205–209 (1989).
4. Pfanner, N. *et al.* *Trends Biochem. Sci.* **21**, 51–52 (1996).
5. Mannella, C. A., Neuwald, A. F. & Lawrence, C. E. *J. Bioenerg. Biomembr.* **28**, 163–169 (1996).
6. Vestweber, D. & Schatz, G. *Nature* **338**, 170–172 (1989).
7. Schnell, D. J. *et al.* *Trends Cell Biol.* **7**, 303–304 (1997).
8. Heins, L. & Soll, J. *Curr. Biol.* **8**, 215–217 (1998).
9. Hinnah, S. C., Hill, K., Wagner, R., Schlicher, T. & Soll, J. *EMBO J.* **16**, 7351–7360 (1997).
10. Crowley, K. S., Liao, S. R., Worrell, V. E., Reinhart, G. D. & Johnson, A. E. *Cell* **78**, 461–471 (1994).
11. Hönlinger, A. *et al.* *EMBO J.* **15**, 2125–2137 (1996).

Active galaxies

Light jets near light speed

Tom Jones

What creates the high-speed jets of material that emerge from quasars, radio galaxies and other active galactic nuclei? On page 457 of this issue¹, Wardle *et al.* report the detection of a subtle but telling feature in the radio emission from the inner jet of the violently variable quasar 3C279. They present the first reliable images and spectral distributions of circularly polarized emission from such a jet — the strongest evidence yet that these jets are mostly made of electron–positron pairs, many at low energies. Similar reports by these observers are available² for three other galaxies, 3C273 (Fig. 1), PKS0530+134 and 3C84. So what does that tell us about the processes that accelerate these jets, presumably in the innermost regions of active galaxies?

Any galaxy whose central emission does not appear to be coming from stars is said to have an active galactic nucleus (AGN). AGNs often emit radiation at a very wide range of wavelengths, from radio waves to γ -rays, and some of them are the most luminous objects in the Universe. Although they are diverse in their spectrum and luminosity, we can explain most of their properties using a single, fairly simple theoretical picture. But this picture has two superficially

contradictory central elements.

We believe that in all these objects, gas falls onto a very massive black hole in the galactic nucleus³ — sometimes called the 'beast'. Because it has some angular momentum, the gas forms an accretion disk^{4,5}. Friction dissipates the angular momentum, allowing the gas to move inwards and heating it up, and thermal emission from this hot gas is thought to be responsible for most of the energy released by a typical AGN,

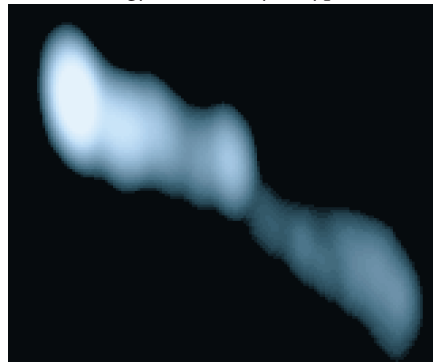


Figure 1 A jet in the quasar 3C273. Its inner few hundred parsecs show up here in 15-GHz radio waves. It is probably an electron–positron plasma, travelling away from a central supermassive black hole at nearly the speed of light.

especially between the infrared and X-ray bands. That much of the picture is fairly clear, both observationally and theoretically. We even find similar accretion disks in other compact astronomical objects, such as newborn stars surrounded by remnants of their birth, and white dwarfs and neutron stars collecting mass from close binary companions, so AGN disks are only the most extreme cases of a common natural phenomenon.

That brings us back to jets, the second element of the AGN picture. Given that AGNs are powered by infall, it is perhaps strange that many have very fast outflows, in the form of highly collimated jets of plasma. These jets are believed to be produced by an interaction between the accretion disk and the central black hole³, in a region so small that it may be impossible to observe directly. So jet properties reflect conditions in a crucial region that is otherwise very difficult to probe.

Since the first jet-like structure was recognized in the relatively nearby galaxy M87 in the 1950s, jets have been studied in many electromagnetic wavebands, and especially in radio waves. Astronomers have discovered that the giant lobes of radio emission seen on either side of some AGNs are connected to the parent galaxy and its nucleus by jets that are sometimes hundreds of kiloparsecs long. Even more dramatically, they have found evidence for rapid expansion along the first few parsecs of some jets⁶, at up to ten or more times the speed of light. At first these 'superluminal motions' caused quite a flurry, because they appear to violate relativity, but it was soon accepted that the apparent superluminal motion is a relativistic illusion that results when the jet is tilted slightly away from the line of sight, and the speed is close to the speed of light (perhaps just a fraction of a per cent less). Within the past few years, similar discoveries have been made for stellar-mass X-ray objects within the Milky Way⁷. So how are jets accelerated to almost the speed of light?

One of the keys to unravelling the origin of jets is to establish their composition. The radio emission is almost certainly synchrotron radiation coming from highly relativistic electrons in a magnetic field. But what balances the electrons' negative charge? Is it protons or positrons? Enter Wardle *et al.*¹. They have managed to detect and image circularly polarized radio emission coming from the parsec-scale jet in 3C279. The observations are an impressive feat, as this polarized signal is very weak and a number of pitfalls must be avoided. Their reward is a clear, direct look at the mass and energy composition of the plasma in the radio jet.

In general terms, polarization depends on the symmetries of the environment where radiation is generated, and through which it passes. To a first approximation, synchrotron emission is plane polarized perpendicular to

the direction of the emission region's magnetic field. But added to that is Faraday rotation, a birefringence effect that occurs in any cool plasma whose opposite charges have different masses, so that one is more sluggish than the other. This causes the plane of polarization to rotate as a function of depth through the plasma, and also as a function of frequency. Radio astronomers commonly use this frequency dependence to estimate the density or the magnetic field strength of the plasma through which synchrotron emission has passed, assuming it to be an equal mixture of electrons and protons, whose masses differ by a factor of about 2,000.

But if the plasma is made of highly relativistic particles, or if the masses of the opposite charges are alike, plane and circular polarizations are interconverted⁸. The same effect (caused by entirely different mechanisms) enables calcite crystals to be used as polarizers on Earth. Furthermore, in a relativistic plasma the circular birefringence has a different frequency signature from that of Faraday rotation. So if both effects are present and measurable, it is possible to determine both the density and the mass composition of the plasma. And as a relativistic particle's effective mass depends on its energy, one can not only distinguish electron–proton from electron–positron plasmas, but also estimate the full energy composition⁹.

In 3C279, Wardle *et al.*¹ have found that the jet is mostly made of electrons and positrons, and that their spectrum must extend to low energies. But, as the results also demand residual Faraday rotation, there must still be some heavier charged particles such as protons, or a situation in which the lowest energies for electrons and positrons are different, changing their effective masses. It is too soon to distinguish between these possibilities, but observations that follow variations in the Faraday rotation and polarization could solve that puzzle. However, even without that information, we have a new clue about the formation of jets.

Jets are probably accelerated either through magnetic effects or through pressure from the intense radiation field near the black hole or the accretion disk. The most obvious composition would be an ordinary electron–proton plasma, similar to the material falling towards the black hole. But, because protons are so massive, it has been hard to build a theoretical model that accelerates proton-loaded outflows up to near light speed. Also, the high-energy γ -radiation in the innermost regions of at least some AGNs would interact with protons to produce mesons in large quantities. Those in turn would decay, leaving behind electrons, positrons and neutrinos. So we already had reason to believe that the plasma should be partly made of electron–positron pairs, but until now the observational case has not been compelling.

There are several ways that electron–positron pairs might be produced in large quantities as the by-products of energetic particle collisions very near the black hole. Most mechanisms fall into one of two classes. One involves electromagnetic cascades, initiated by the collision of two photons possessing enough energy to produce an electron–positron pair, each of which then emits photons that can repeat the process. This possibility is especially likely within a few black-hole radii of the hole itself, where the radiation field is most intense. But if a pair plasma is produced too close to the black hole in the presence of a luminous accretion disk, radiation drag on the outflowing plasma will prevent it from reaching the speeds seen in radio jets, and the electrons and positrons will cool rapidly.

The other possibility involves the decay of mesons produced when a very energetic proton collides with another proton or a photon. That would require acceleration of protons to high energy very close to the black hole, perhaps by shocks. Such models may

get around the radiation-drag issue. They may also have advantages made apparent by the new data: meson-decay models could naturally include a residual component of ordinary proton–electron plasma, or accommodate different proportions of electrons and positrons at the lowest energies. So these new observations may have helped to nail down the mechanisms for jet formation in AGNs. □

Tom Jones is in the School of Physics and Astronomy, University of Minnesota, 116 Church Street SE, Minneapolis, Minnesota 55455, USA.
e-mail: twj@ast1.spa.umn.edu

1. Wardle, J. F. C., Homan, D. C., Ojha, R. & Roberts, D. H. *Nature* **395**, 457–461 (1998).
2. Homan, D. C., Wardle, J. F. C., Roberts, D. H. & Ojha, R. (personal communication).
3. Begelman, M. C., Blandford, R. D. & Rees, M. J. *Rev. Mod. Phys.* **56**, 255–351 (1984).
4. Tanaka, Y. *et al.* *Nature* **375**, 659–661 (1995).
5. Nakai, N., Inoue, M. & Miyoshi, M. *Nature* **361**, 45–47 (1993).
6. Zensus, J. A. *Annu. Rev. Astron. Astrophys.* **35**, 607–636 (1997).
7. Mirabel, I. F. & Rodríguez, L. F. *Nature* **371**, 46–48 (1994).
8. Jones, T. W. & O'Dell, S. L. *Astrophys. J.* **214**, 522–539 (1977); **215**, 236 (1977).
9. Jones, T. W. *Astrophys. J.* **332**, 678–695 (1988).

Immunology

Sinking surveillance's flagship

George Klein and Eva Klein

Epstein–Barr virus (EBV; Fig. 1) is the most potent transforming and potentially tumorigenic virus known. Its transforming power is manifested in its ability to change human B lymphocytes into 'immunoblasts', activated immunoglobulin-producing cells that contain the viral genome and can divide continuously. That at least 90% of all human beings live in peaceful equilibrium with EBV for their entire life is an impressive testimony to immunological surveillance, for the proliferation of EBV-infected cells is kept under strict control by the T cells of the immune system. Disease — the most dramatic example being X-linked lymphoproliferative syndrome, XLP — occurs only as a biological accident. In *Nature Genetics*¹ and on page 462 of this issue² come reports of the cloning and characterization of the XLP gene. Unexpected-

edly, it is mutation of a small protein involved in T-cell regulation that is responsible for the breach in control of EBV-transformed immunoblasts — for sinking this flagship of immune surveillance.

Highly transforming viruses such as EBV that infect most individuals of their host species have selected their hosts for the virtually watertight protection they provide against the development of virally driven tumours, while permitting the survival and propagation of the virus. EBV and its close relatives are found in all Old World (but not New World) primates; so coexistence of virus and host is of ancient standing, and it is safeguarded at many levels. In healthy EBV carriers, the virus hides largely, if not exclusively, in the immune system, particularly in B cells, which are also its main transformation target. Equilibrium is maintained by a

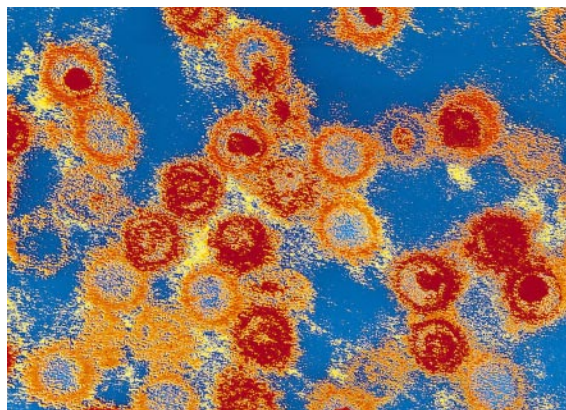


Figure 1 Particles of Epstein–Barr virus, $\times 58,000$. The virus was first isolated in 1964, by Epstein and Barr, in cell cultures from lymphoma tissue.

finely poised balance between the regulation of viral gene expression and the immune system, particularly T cells (see ref. 3 for review).

That equilibrium can be upset by infections (for instance with HIV), or clinical or inherited impairment of the immune system. This may lead to the fatal form of infectious mononucleosis (glandular fever); or to progressively growing immunoblast cells that may turn into monoclonal lymphoma, malignant tumours of the lymph nodes. But immune control of EBV-transformed immunoblasts is remarkably robust. For example, the immune systems of transplant patients are strongly suppressed by drugs, but only a minority of recipients develop lymphoproliferative disease; even then they can sometimes be successfully treated by the infusion of appropriate T cells.

Inherited immunodeficiencies that allow lymphoproliferative disease to occur may point to the Achilles heel of the host defence system. The X-linked lymphoproliferative syndrome, described by David Purtilo in 1969, is the most prominent of these diseases and has alternative EBV-induced outcomes^{4,5}. The syndrome manifests itself from 2.5 years of age and onwards, depending on the time of EBV infection, and all sufferers die by the time they are 40. Male carriers of the XLP trait respond to primary EBV infection with uncontrolled proliferation of both T and B cells. Fatal infectious mononucleosis occurs in 51% of patients; acquired hypogammaglobulinaemia (immunoglobulin

deficiency) in 31%; and malignant lymphoma alone or in combination with the other symptoms in 26%.

Almost a decade ago genetic linkage research localized the *XLP* locus to region 24–25 on the long arm of the X chromosome (Xq24–25)⁶. Subsequently, the critical region was narrowed down considerably^{7,8}, but attempts to clone the gene failed. Now, however, two groups have done so.

Coffey *et al.*¹ report the positional cloning of the *XLP* gene, which they call *SH2D1A*. It has four exons and encodes a hitherto unknown protein, of 128 amino acids, that contains a single SH2 (src homology 2) domain; this is a protein module that specifically binds to tyrosyl phosphorylated peptides on signalling proteins. *SH2D1A* is expressed mainly in thymus and lung and in all lymphocyte populations assayed. Coffey *et al.* found that it was mutated in 9 out of 16 unrelated XLP patients, but not in male family members of those patients and other controls unaffected by the disease.

Sayos *et al.*² have reached the same goal but by a different route. They studied SLAM (signalling lymphocyte activation molecule), a glycosylated transmembrane protein also known as CDw150. SLAM has a high affinity for itself and is centrally involved in the bi-directional stimulation of T and B cells⁹. When activated, it mediates expansion of activated T cells during immune responses¹⁰, induces production of interferon- γ and changes the functional profile of subsets of T cells. Signalling through SLAM–SLAM

binding during mutual interaction between B cells, and between B cells and T cells, increases the expansion and differentiation of activated B cells¹¹.

Sayos *et al.* used the human SLAM cytoplasmic domain as bait in the yeast two-hybrid system. From a human T-cell library, they isolated eight DNA clones that encoded a 128-amino-acid polypeptide, designated SLAM-associated protein (SAP). As well as a single SH2 domain, this protein has a short tail, 26 amino acids in length, that is unlike any known protein domain. SAP is expressed in T cells (but not in B cells), both in humans and in mice. It turns out that SLAM and SAP interact in T cells through SAP's SH2 domain.

Using a clone that contained all four exons of mouse SAP, Sayos *et al.* localized the gene to part of the mouse X chromosome corresponding to human Xq25. To test the possibility that SAP was encoded by the *XLP* gene, SAP complementary DNAs were isolated from the blood cells of three XLP patients. All three cDNAs were mutated or deleted, in contrast to those from a healthy brother and other controls. The mutant SAPs did not bind SLAM.

In thinking how the mutation of SAP can account for the XLP diseases, Sayos *et al.* suggest that SAP is a natural inhibitor of SLAM. Consistent with this idea, they found that SAP binding blocks the recruitment of the tyrosine phosphatase SHP2 to the phosphorylated cytoplasmic domain of SLAM. Upon T-cell activation, SLAM may switch from a

Motor proteins

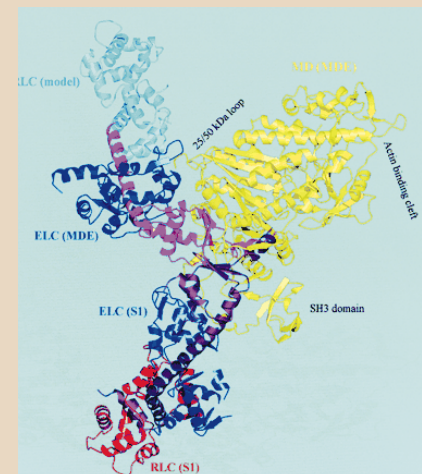
Sighting of the swinging lever arm of muscle

The first sight of the structure of muscle's power stroke is shown in this figure, taken from a paper by Carolyn Cohen and colleagues published on 4 September (*Cell* 94, 559–571; 1998). The authors report the crystallographic structure, at around 3 Å resolution, of the motor domain (MD) of myosin from the thick filaments of smooth (chicken gizzard) muscle in complex with its essential light chain (ELC). This complex (MDE) forms the main part of the myosin head, which converts chemical to mechanical energy via a conformational change induced by the hydrolysis of ATP (catalysed by actin in the thin filaments). Understanding the mechanism by which myosin (and other motor proteins) converts chemical to mechanical force — manifested as sliding between the thick and thin filaments of muscle fibres — is a long-sought goal, and the new data provide an exciting glimpse of the structural basis of this mechanism.

The tail of the myosin head contains a very long α -helical region, thought to be a

'lever arm' that rotates when ATP binds and is hydrolysed — this rotation is the basis of the power stroke. The crystallographic structure of the end of the power stroke has been described previously, and there have been indications that there is a swinging lever arm. But until the new results of Cohen and colleagues, which show the beginning of the power stroke, there has been no unequivocal structural evidence for this hypothesis.

The figure consists of two superimposed structures. The motor domain of myosin is in yellow, and the magenta and pink regions of the molecule are the lever arm. The bottom structure represents the end of the power stroke, in which the molecule is free of bound nucleotide. The top structure represents the beginning of the power stroke, in which an ATP analogue is bound to the myosin. In this structure, the light-blue region of the molecule is a model, necessary because the construct used lacks



a regulatory light chain (RLC). The angle between the ends of the two lever-arm conformations is about 70°, which would produce a filament displacement of about 10 Å. This distance is consistent with measurements obtained from functional studies of muscle fibres, and from *in vitro* motility assays of the component proteins.

Maxine Clarke

NATURE

100 YEARS AGO

Recent researches by Surgeon-Major Ronald Ross have shown that the mosquito may be the host of parasites of the type of that which causes human malaria. Ross has distinctly proved that malaria can be acquired by the bite of a mosquito, and the results of his observations have a direct bearing on the propagation of the disease in man. Dr. P. Manson describes the investigations in a paper in the *British Medical Journal*, and sums them up as follows: – The observations tend to the conclusion that the malaria parasite is for the most part a parasite of insects; that it is only an accidental and occasional visitor to man; that not all mosquitos are capable of subserving it; that particular species of malaria parasites demand particular species of mosquitos; that in this circumstance we have at least a partial explanation of the apparent vagaries of the distribution of the varieties of malaria. When the whole story has been completed, as it surely will be at no distant date, in virtue of the new knowledge thus acquired, we shall be able to indicate a prophylaxis for malaria of a practical character, and one which may enable the European to live in climates now rendered deadly by this pest.

From *Nature* 29 September 1898.

50 YEARS AGO

"Thomas Jefferson Among the Arts" – Thus we find Jefferson the revolutionary, workman, writer, thinker, toiling ever and anon to establish culture, both abstract and material, in his youthful America. To him there was no difference between 'pure' and 'applied' (whether art or science); all was for the benefit of mankind, and the pursuit of the good. Naturally, this led to some strong likes and dislikes, short shrift to Plato and Samuel Johnson, to mention but two. Love of contrast is characteristic; formal Palladian architecture, surrounded by 'serpentine' gardens, as if to say "Oh Western Wind, blow soft and kind" upon the exceedingly solid and uncompromising stone buildings he admired and advocated. The author of the Declaration of Independence is entitled to such light and shade in his dealings with contemporary events.

From *Nature* 2 October 1948.

signalling cascade that is dependent on SAP (and probably on fyn, a member of the src tyrosine kinase family) to one that depends on SHP2. Put more generally, Sayos *et al.* propose that SAP controls the signal-transduction pathways initiated by interactions between SLAM molecules at the interface between T and B cells.

Why should the impairment of SAP, an inhibitor of SLAM-triggered T-cell activation, inhibit the T-cell response against EBV-transformed immunoblasts? The answer will come from clarification of the SLAM pathway, of the interactions between T cells and EBV-infected B cells, and of SAP-dependent T-cell inhibition. EBV-transformed B cells express SLAM at a high level, and are also professional antigen-presenting cells; that is, they stimulate T cells to proliferate and generate both CD4⁺ and CD8⁺ effectors targeted against EBV proteins involved in growth transformation. Even a seemingly minor disturbance in SAP may perturb the finely tuned interactions between the different effectors, with disastrous consequences. Eventual apoptosis of the activated, but SAP-defective and therefore dysregulated EBV-specific T cells, is one of the obvious possibilities.

Epstein-Barr virus is unique in being both an inhabitant of B cells and an active player in the B-cell system, especially in the interactions between the EBV-infected immunoblast cells and T cells. That might

explain the curious fact that XLP patients have a selective defect in their T-cell response to these immunoblasts but not to cells infected with other viruses, and that they seem to be more prone to EBV-associated disease than are patients with other inherited immune deficiencies. Conceivably, regulation of the SAP-SLAM pathway may also differ with age. If so, it might help explain why primary EBV infection remains silent in most small children and usually causes mononucleosis only in adolescents and young adults.

The identification of the XLP gene, alias *SH2D1A* or *SAP*, and its signalling function opens the way to a better understanding of the highly sophisticated immune response against EBV-transformed cells. It may improve the diagnosis of a severe but enigmatic disease with variable manifestations, and it may also in time lead to treatments. □

George Klein and Eva Klein are at the Microbiology and Tumor Biology Center, Karolinska Institute, PO Box 280, S-171 77 Stockholm, Sweden.
e-mail: Georg.Klein@mtc.ki.se

1. Coffey, A. J. *et al.* *Nature Genet.* **20**, 129–135 (1998).
2. Sayos, J. *et al.* *Nature* **395**, 462–469 (1998).
3. Klein, G. *Cell* **77**, 791–793 (1994).
4. Purtilo, D. T. *Adv. Cancer Res.* **34**, 279–312 (1981).
5. Seemayer, T. A. *et al.* *Pediatr. Res.* **38**, 471–478 (1995).
6. Skare, J. C. *et al.* *Human Genet.* **82**, 354–358 (1989).
7. Lanyi, A. *et al.* *Genomics* **39**, 55–65 (1997).
8. Lamartine, L. *et al.* *Eur. J. Hum. Genet.* **4**, 342–351 (1996).
9. Cocks, B. G. *et al.* *Nature* **376**, 260–263 (1995).
10. Aversa, S. *et al.* *J. Immunol.* **158**, 4036–4044 (1997).
11. Punnonen, J. *et al.* *J. Exp. Med.* **185**, 993–1004 (1997).

Neurobiology

See and grasp

Nikos K. Logothetis

To be able to grasp and manipulate objects, we need to know not only their location, but also their shape, orientation and size. Visual information about such properties determines the anticipatory posture of the hand and fingers during reaching, and is critical for controlling skilled actions such as precision grip¹. But the neural representations of shapes may be different from the representations that mediate the visual control of action. For example, although object recognition relies on representations that do not vary with changes in orientation, location and size, to act on these objects we probably need representations that are sensitive to just such changes.

On page 500 of this issue, Sereno and Maunsell² address the questions of how object- and space-related information is integrated, and how the brain uses the invariant representations that underlie visual perception to guide our geometry-specific grasping. In primates, the perception of objects and space seems to be mediated by two neural pathways, which are functionally distinct and anatomically segregated^{3–5}

(Fig. 1). Although both emanate from the primary visual cortex, one then stretches to the temporal lobe (ventral pathway), and the other to the parietal lobe (dorsal pathway)³. Sereno and Maunsell now suggest that there may be visual shape representations in both of these visual pathways, some directly suitable for visually guided action. Recording from the posterior parietal cortex of monkeys — an area thought to be involved in spatial vision — they find that neurons in the lateral intraparietal area, one of the functional areas of the parietal cortex, respond selectively to images of simple, two-dimensional shapes.

Previous studies^{6,7} showed that neurons in the parietal lobe of monkeys are sensitive to shape. Monkeys were trained to reach and grasp real, three-dimensional objects, either immediately after the object was presented or after a delay period of several seconds during which the object could not be seen. Posterior parietal neurons were found to be selective to both the visual appearance of simple, geometrical solids, and the monkey's short-term memory of them. Moreover, most of these cells showed shape-

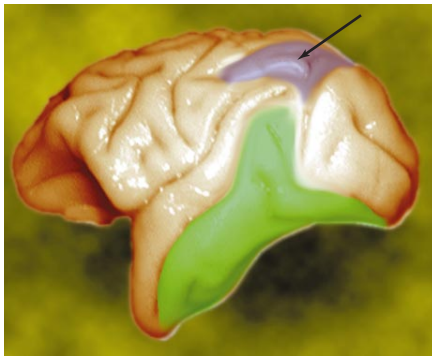


Figure 1 Lateral view of the macaque brain. The purple and green areas show cortical regions that are involved in processing spatial and object information, respectively. The area reported by Sereno and Maunsell lies within the intraparietal sulcus, shown by the arrow.

selective activity, even when the monkey was required to refrain from grasping the object. Nonetheless, in all of these experiments the monkeys were trained to manipulate the test objects and, presumably, had developed the representations required for their hand manipulation through both visual and somatosensory information.

What makes the findings of Sereno and Maunsell remarkable is that their monkeys were not involved in a hand-manipulation task, and they could not touch the stimuli (two-dimensional images of objects). In fact, the animals were trained only to perform a simple fixation task or a visual delayed match-to-sample task. During the fixation task, the monkeys looked at a screen, holding their gaze on a central spot, while a shape was presented within the receptive field of the recorded neuron. During the delayed match-to-sample task, the animals fixated a central spot and, shortly afterwards, a sample shape was superimposed on the spot. This was then replaced by three shapes equidistant from the fixation spot, and the monkeys had to make an eye movement to the test shape that matched the sample shape for a juice reward. Surprisingly, many of the recorded parietal neurons — just like the temporal or prefrontal cells reported by other investigators — showed differences in activity as different shapes were presented during the fixation task. Some also showed shape selectivity during the delay period of the match-to-sample task.

Why should shape be represented in the posterior parietal cortex? Our perception of shapes is probably mediated by neurons in the temporal lobe. Although lesions in this area of the brain severely interfere with pattern perception and recognition, lesions in the parietal lobe only affect spatial vision (our sense of where things are), leaving pattern vision intact³. Most neurons in the temporal lobe respond selectively to simple or complex visual patterns, including views of human and monkey faces, indicating that

this region is critical in shape perception. Shape selectivity is also found in areas of the prefrontal cortex. These are interconnected with the visual areas of the temporal lobe, and are thought to mediate the working memory for visual objects⁸.

Neurons in the temporal lobe tend to be position and size invariant; that is, they lack properties that are critical for manipulating objects by hand. Although no data are available, shape selectivity in the posterior parietal cortex may turn out to be specific for the size range of objects that an animal could possibly manipulate. Shape selectivity in the parietal cortex may also be specific for orientation with respect to a reference frame centred on the viewer or on some other object — a property that is rarely seen in the responses of temporal neurons. On the other hand, selectivity to material properties of objects, such as colour and textures, would be meaningless for a system that underlies hand-manipulation of objects, but essential for certain recognition tasks. So the existence of separate temporal and parietal shape representations may be partly due to the different output requirements of the visual system, as neuropsychological studies also suggest⁹.

Finally, based on the neuronal properties of the posterior parietal cortex, it has been suggested that this region of the brain is the main area that mediates visuospatial attention^{10,11} or, alternatively, an animal's intention to reach a particular point in space¹². The data presented by Sereno and Maunsell indicate that at least some areas of the parietal cortex may be important for switching attention or instigating an intentional movement, not only to particular locations, but also to particular objects that are targets for either action or identification. Shape- and location-triggered attentional or intentional shifts are bound to require neurons that discriminate shapes to some extent. □

Nikos K. Logothetis is at the Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, D-72076 Tübingen, Germany.
e-mail: nikos.logothetis@tuebingen.mpg.de

1. Jeannerod, M. *The Neural and Behavioural Organization of Goal-Directed Movements* (Oxford Univ. Press, 1988).
2. Sereno, A. B. & Maunsell, J. H. R. *Nature* **395**, 500–503 (1998).
3. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (ed. Ingle, D. J.) 549–586 (MIT Press, Cambridge, MA, 1982).
4. Logothetis, N. K. & Sheinberg, D. L. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
5. Colby, C. L. *Neuron* **20**, 15–24 (1998).
6. Taira, M., Mine, S., Georgopoulos, A. P., Murata, A. & Sakata, H. *Cerebral Cortex* **5**, 429–438 (1990).
7. Murata, A., Gallese, V., Kaseda, M. & Sakata, H. *J. Neurophysiol.* **75**, 2180–2186 (1996).
8. O'Scalaidhe, S. P., Wilson, F. A. W. & Goldman-Rakic, P. S. *Science* **278**, 1135–1138 (1997).
9. Goodale, M. A. & Milner, A. D. *Trends Neurosci.* **15**, 20–25 (1992).
10. Colby, C. L. *J. Child. Neurol.* **6**, 90–118 (1991).
11. Corbetta, M., Miezin, F. M., Shulman, G. L. & Petersen, S. E. *J. Neurosci.* **13**, 1202–1206 (1993).
12. Snyder, L. H., Batista, A. P. & Andersen, R. A. *Nature* **386**, 167–170 (1997).

Daedalus

Psychic misperceptions

In Everett's 'many worlds' view of quantum mechanics, the complete multidimensional Universe can be divided into many vector subspaces. Each subspace is a physically real 'parallel world'. Quantum-probability paradoxes neatly disappear. Schrödinger's famous cat, whose life hangs on one random quantum event, lives in some of the Everett worlds, but dies in others.

Clever, says Daedalus; but not clever enough. Why should the Universe divide so neatly into perfect physically real worlds? By choosing a different set of orthogonal vectors, you could divide it just as validly into worlds whose quantum states were hopelessly mixed. Their inhabitants would be like Schrödinger's cat before it is observed: neither alive nor dead, but in a ghostly, non-physical, quantum superposition of these states.

So Daedalus sees the complete Universe as a set of physically real Everett worlds, embedded in a matrix of quantally mixed, physically non-real, ghostly worlds. He identifies this mystic matrix with the spiritual world of ghosts, telepathy, and so on. Indeed, it may act as a telepathic channel for messages from other Everett worlds. This theory explains the deplorably unreliable nature of mystic insights, telepathic intuitions and so on. They may in fact refer to some other world.

But how to tell? Daedalus reckons that a complementarity principle must apply. To sustain the correct probabilities of its quantum states, a live Schrödinger cat in one world implies a dead one in another; a successful lucky chance in one world must fail in another. Daedalus recalls a study of intuition in business executives. Successful ones scored significantly better than chance; but failing ones scored worse than chance. Clearly, says Daedalus, these perverse 'anti-psychics' were, sadly for them, tuned to another world. So Daedalus plans to seek out such rare and gifted individuals among bankrupt businessmen, failed spiritualists and inspired losers of all kinds. He will then look for statistical agreements among their hopeless fantasies. Tantalizing glimpses of some other Everett world may emerge.

Although this world will probably be very close to ours in the universal 'holospace', Daedalus can see no way of exchanging material as well as ideas. He does wonder, however, whether its inhabitants find extra socks in the wash, but suffer mysterious losses of wire coat-hangers.

David Jones

Obituary

Thomas Kreis (1952–98)

Cell biologist

Among the 229 lives claimed by the crash of the Swissair flight 111 on 3 September 1998 was that of Thomas Kreis, one of the world's leading cell biologists. At 46, he was at the peak of his life and his career — a new father, professor and newly appointed chairman of the department of cell biology at the University of Geneva.

Kreis entered cell biology in the 1970s at a time when the discipline was changing radically. Until then, it had mainly involved the morphological description of cells and tissues. With the advent of molecular biology, a small but influential group initiated analysis of the molecular mechanisms responsible for cellular organization by tackling the problem of sorting: how do the thousands of proteins made by each cell find their correct intracellular and extracellular destinations?

Biochemical assays were devised for studying how proteins synthesized by the protein factories in the cytoplasm are directed to different cellular compartments such as the endoplasmic reticulum, Golgi complex, endosomes, lysosomes and plasma membrane. At the same time, cell biologists started to unravel the functions of the various compartments themselves. But the overall picture was static. It was clear that different compartments had to communicate with each other. But how this is done, and how organelles such as the Golgi complex are built and maintained, was unknown. The biochemical and molecular biological approaches to cell biology were inherently reductionist, and of necessity often destroyed cellular context.

It was here that Kreis came into the picture. He was one of the first to set out to study cellular function in the context of living cells. During his PhD thesis at the ETH in Zurich, 1978–81, he was already microinjecting fluorescently labelled proteins into live cells and following their fate by fluorescence microscopy. This technique was taken further during his postdoctoral year with Joseph Schlessinger and Benjamin Geiger at the Weizmann Institute in Israel. There he continued his analysis of cytoskeletal elements in the cell, the dynamic network of proteins responsible for cell movement and intracellular transport. This was at a time when the war in Lebanon was at its fiercest and his colleagues wondered, sometimes not happily, at his ability to concentrate on his work ("You are forgetting that I am



Swiss, and therefore neutral!"). Kreis then spent 1982–83 at the Massachusetts Institute of Technology with Harvey Lodish, a year that influenced him greatly and during which he showed that microinjection could be adapted to probe a wide variety of problems in protein transport.

In 1983, Kreis became a group leader in the cell biology programme at the European Molecular Biology Laboratory, Heidelberg; this was when his science really hit its stride and his reputation became established. Nine years later, at the age of 40, he moved to Geneva as a professor.

During this 15-year period, Kreis and his group, together with an ever-growing array of international collaborators, established an information base that helped to provide much of our current insight into how intracellular organelles arrange themselves on microtubule tracks and how these tracks are used to transport parcels among compartments in the cell. He was, for instance, the first to establish links between the containers and the microtubules by discovering proteins such as CLIP-170, which connects endosomes to microtubules and so organizes their movements throughout the cell.

In these studies, Kreis might have been inspired by his skill as an expert skier. In one of his last papers, he and his group showed just how intricate the functioning of the cellular transport pathways could be. On the ski slopes, skiers without valid passes sometimes sneak into the gondolas; in the endoplasmic reticulum, proteins not intended for transport to the Golgi complex sometimes sneak into transport

vesicles along with proteins actually intended for secretion.

Kreis's imaging of these events in living cells, through the use of reporter molecules labelled with green fluorescent protein, showed that the interlopers are removed from the moving gondolas and packaged into other vesicles that return them to the endoplasmic reticulum. Organizing the entire process is the dynamic network of microtubule cables on which these vesicles move. But what organizes the cables themselves remains a mystery: after depolymerization, microtubules always contrive to reassemble in almost exactly the same pattern.

Kreis's legacy to science extends beyond his discoveries in cell biology to the example set by his style. His approach was characterized by its straightforwardness. As a ski instructor, he used to start his mornings from the top of the highest mountain and then take the straightest route down to the bottom to warm up. Likewise, in research he was always looking for short cuts and quicker ways to the goal; and so, in his dealings with collaborators and colleagues he was honest yet always supportive.

Another element of his style was the understated grace of his work. His beautiful fluorescence micrographs and sensitive hand-drawn images are reflections of a man with a poetic streak (seen also in his extraordinary collection of roses and cactuses).

Kreis also did much for the common good in organizing advanced courses and meetings. Most notably he was one of the initiators of an organization called the European Life Scientist Organization (ELSO). This will start its activities in Geneva in September 2000 with a large congress planned for 3,000–4,000 participants. The hope is that ELSO will become a general forum for catching the excitement in the molecular life sciences and help to raise the standards of biomedical research all over Europe.

Thomas Kreis leaves a wife, son and stepdaughter. He will be missed most by them, but also by the whole community of cell biologists. We have lost, far too prematurely, one of our pioneers.

Kai Simons and Ira Mellman

Kai Simons is in the Cell Biology Programme, European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, 69112 Heidelberg, Germany.

e-mail: simons@embl-heidelberg.de

Ira Mellman is in the Department of Cell Biology, Yale Medical School, 333 Cedar Street, PO Box 208002, New Haven, Connecticut 06510-8002, USA.

e-mail: ira.mellman@yale.edu

Haeckel's hierarchies

The now-familiar concept of drawing lines of descent as trees seems harmless enough. But it led Ernst Haeckel (who bent the evidence to prove his Darwinian theories) into believing in the evolution of a Germanic super-race.

Martin Kemp

The tree as a schema for lines of descent is one of those devices that have become so familiar as to give us little pause for thought. Yet it is often rich in implicit meaning, far beyond its immediate graphic utility.

The leading designer of evolutionary trees was the fervent German Darwinian, Ernst Haeckel. Famed for his beautifully illustrated publications of the Radiolaria — those geometrical masterpieces of micro-engineering — and as an anti-Catholic polemicist, Haeckel sought to so extend and consolidate Darwin's theory that it would brook no contradiction. He insisted that "ontogeny is a short and rapid recapitulation of phylogeny" — to the extent of bending his visual evidence to demonstrate identical stages in the embryological development of different species.

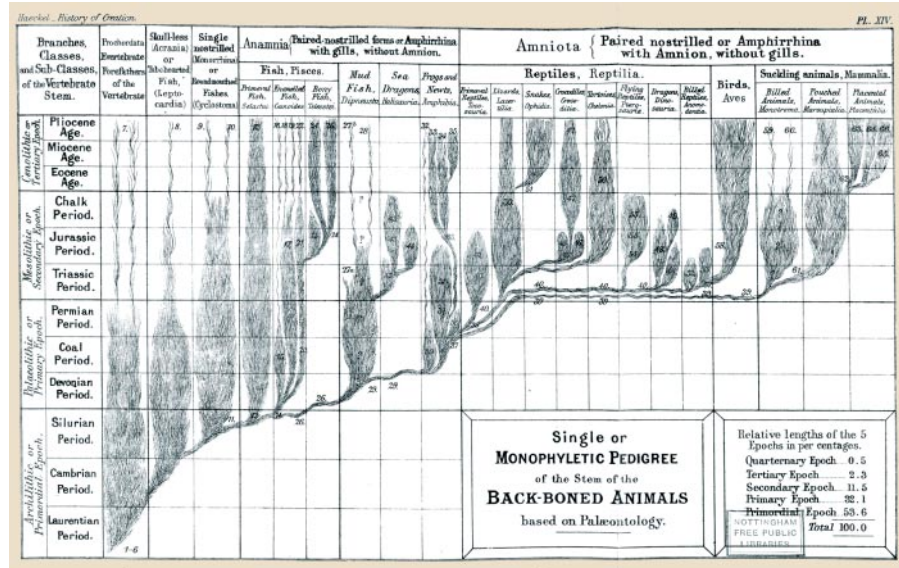
For Haeckel, every act of drawing was simultaneously an act of observation, analysis and demonstration in which appearance and interpretation were blended.

Haeckel aspired to a monist philosophy which dissolved the traditional dualisms of science and religion, reason and revelation, soul and body, mind and matter, man and animals, living and non-living, and organic and inorganic. His monism, unifying on the surface, was nevertheless suffused with hierarchical values, not least in his trees.

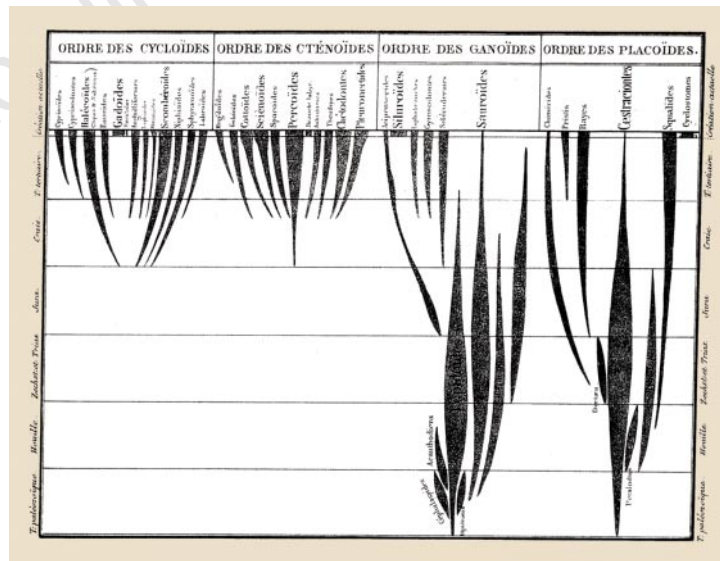
In his first major contribution to evolutionary theory, his *General Morphology of Organisms* in 1866, the tree assumes its most organic guise, with nicely rounded trunk and convincing ramifications, while in *The History of Creation* (translated by E. Ray Lankester in 1875–6) the trees are transformed into fern-like structures, with a degree of impressionistic suggestiveness commensurate with the uncertainties of the fossil record.

Such trees were not the inevitable outcome of an awareness of the successive phases of the appearance of species. Louis Agassiz's elegant spindle diagram in his book on fossil fish in 1833, eloquently expounded by Stephen Jay Gould, resists the joining of the branches to the trunks, because Agassiz was "convicted that they do not descend, one from the other". If Agassiz's scheme is concerned with special creations and descent, Haeckel deals with the remorseless ascent of humankind.

Acknowledging that "man is separated from the other animals by quantitative, not



Ernst Haeckel, 'Single or Monophyletic Pedigree of the Stem of the Back-Boned Animals based on Palaeontology' from *The History of Creation*, 1875–6, vol. II, opposite page 233.



Louis Agassiz, 'Diagram of the Evolution of Fishes' from *Les poissons fossiles*, 1833, vol. I, p. 170 (in S.J. Gould, *Eight Little Piggies*, p. 431).

qualitative, differences", Haeckel recognized the continuity of all natural things. However, he argued that:

"if one must draw a sharp boundary, it has to be drawn between the most highly developed and civilised man on one hand, and the rudest savages on the other, and the latter have to be classed with the animals".

Unhappily, on the eve of the First World War, he was inclined to agree with the physicist Wilhelm Ostwald that the Germanic race was leading the way into new evolutionary uplands, having "discovered the factor of organisation", whereas "other peoples live

under the regime of individualism".

Haeckel's hierarchical trees all too readily lent themselves to such forms of social Darwinism and eugenic perfection. However, it would have behoven him — and it certainly repays us — to reflect that the unbroken persistence of some 'primitive' species from the lowest ramifications of his trees can just as readily suggest that the human race, as a newcomer on the evolutionary stage, has a long way to go to demonstrate comparably sustained durability.

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

Pollen analysis reveals murder season

Pollen analysis is well established in palaeogeographical research^{1,2}, but it remains an underused resource in forensic science^{3,4}. So far it has been used to investigate only one reported case of murder⁵, but has not to our knowledge been used to determine the time of death. We examined a common grave containing 32 murder victims, and used pollen analysis to establish that the murders must have occurred in the summer of an unknown year. A knowledge of the season of the murder could help to determine who was responsible for the murders.

In February 1994 a common grave containing 32 male skeletons was discovered in Magdeburg in Germany. The area had been under Gestapo control until 1945 when the Soviet occupying forces took over. The identity of both victims and murderers were unclear. Before our pollen analysis, two hypotheses had been proposed: (1) the victims were killed by the Gestapo in the spring of 1945 at the end of the Second World War; and (2) the victims were Soviet soldiers killed by the Soviet secret police following the revolt in the German Democratic Republic on 17 June 1953.

It would have been most unusual for the Gestapo to establish a mass grave in the

centre of a city. According to the few eyewitness reports, the Soviets supposedly interrogated, tortured and executed German and Russian spies in a prison in this area. It has also been reported that Soviet soldiers in Magdeburg were executed after refusing to help break up the German revolt in 1953.

Erdtmann⁵ and Faegri⁶ developed the method of pollen analysis that formed the basis for the reliable determination of species. We used this method to establish the time of death by means of a pollen spectrum with seasonal indication. Air contains pollen throughout the vegetation period, but pollen from a particular species is released only at its flowering time⁷. The pollen is inhaled during breathing, so pollen deposition in the nasal-pharyngeal tract of a corpse can indicate the seasonal spectrum of pollen species at the time of death.

This relationship between pollen accumulation in the human respiratory tract and seasonally occurring pollen species can be clearly demonstrated in a living person. A handkerchief analysis confirmed our assumption: the composition of pollen species contained in the handkerchiefs was exactly what we expected from the pollen occurrence in that season⁷ (Fig. 1). In the

spring, for example, the predominant primary species are alder (*Alnus*), hazelnut (*Corylus*), willow (*Salix*) and juniper (*Juniperus*), followed by combinations of birch (*Betula*), yew (*Taxus*) and poplar (*Populus*). In late June, predominant pollen species typically included various grasses and weeds, plantain (*Plantago*), goosefoot (*Chenopodium*), sagebrush (*Artemisia*) and lime (*Thilia*). Wind-borne pollen species such as birch, pine and hazelnut, which are emitted in extremely large amounts at their flowering time⁸, also occur in small amounts at other times. They therefore represent a seasonally indistinct contamination factor outside the flowering period.

We performed pollen analysis on 21 skulls from the mass grave at Magdeburg. Rinsing fluids of the nasal cavities of seven skulls contained a high content of pollen species that are typical of the summer (see Supplementary information). In particular, we found high concentrations of plantain pollen, a marker with a strong emission peak in June and July, and detected a few other pollen grains from the same emission period, including lime tree and rye. Other pollen species detected at lower levels are not indicative enough of a specific time when considered separately, but support the assumption that death occurred around June or July when considered in combination with the other pollen found. The rinsing liquid also contained fairly high amounts of birch pollen, indicative of spring, but this does not contradict our conclusion because the handkerchief analysis revealed a moderate presence of birch pollen throughout the year.

To ensure that the pollen in the rinsing fluid had been inhaled, and not brought in with the soil, the amounts of pollen in the nasal cavities were compared with those in the cavities of the os sacris. In all cases, the amounts of pollen in the latter were found to be negligible. We concluded that some of the victims had inhaled large amounts of summer pollen shortly before their death. Thus, pollen analysis supported the hypothesis that the grave contained the remains of Soviet soldiers killed by the Soviet secret police following the revolt in June 1953.

Furthermore, anthropometric and DNA data⁹ show that the victims were all young males. The regular layers in which the 32 victims were packed indicates a series of murders within a short period of time. Although some skulls exhibited decayed teeth, there was no sign of any dental work, which is unusual for a population in central Europe in the middle of the twentieth century. Taken together with the results of our

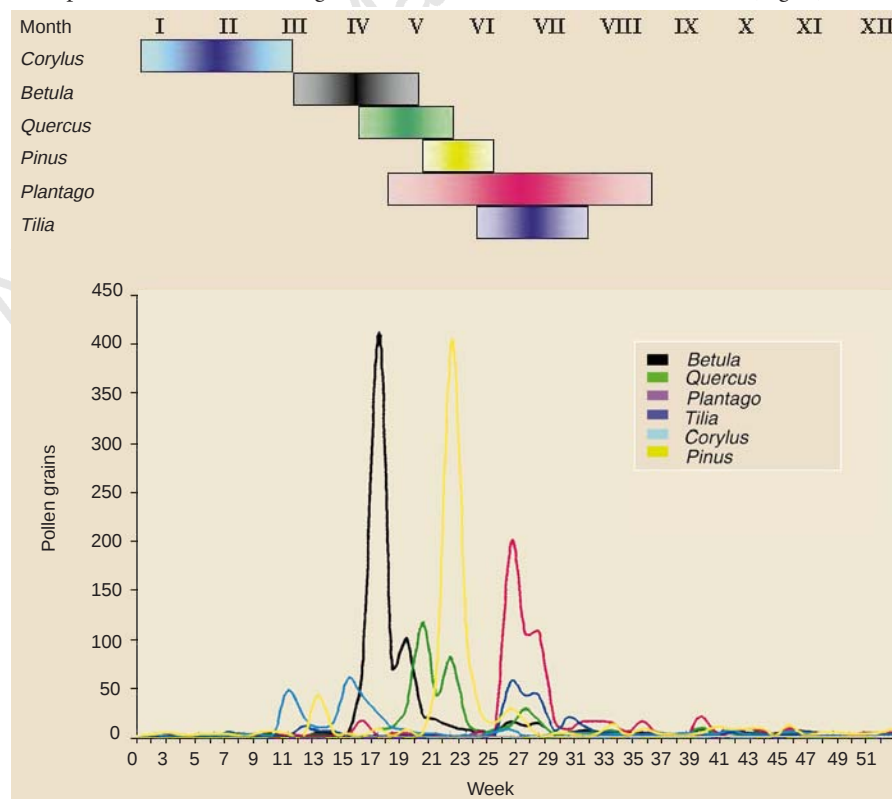


Figure 1 Occurrence of pollen of several species in handkerchiefs as a function of the calendar week (bottom) and in comparison with the pollen flight times in the pollen calendar (top). Data are the number of pollen grains obtained from four separate handkerchiefs. The number of pollen grains was divided by 3 for *Betula* and multiplied by 5 for *Tilia* for ease of data presentation.

pollen analysis, these findings support our conclusion that the murder victims were Soviet soldiers.

R. Szibor*, C. Schubert*, R. Schöning*,
D. Krause*, U. Wendt†

*Institut für Rechtsmedizin,
Otto-von-Guericke-Universität Magdeburg,
Leipziger Strasse 44,
D-39120 Magdeburg, Germany

†Institut für Werkstofftechnik und Werkzeugprüfung,
Otto-von-Guericke-Universität Magdeburg,
Grosse Steinernetischstrasse,
D-39104 Magdeburg, Germany

1. Markgraf, V. *Palynology* 7, 43–70 (1983).
2. Jones, J. G. *Palynology* 18, 205–212 (1994).
3. Bock, H. J. & Norries, D. O. *J. Forensic Sci.* 42, 364–367 (1997).
4. Bryant, V. M., Jones, G. J. & Mildenhall, D. C. *Palynology* 14, 193–208 (1990).
5. Erdtmann, G. *Handbook of Palynology* (Hafner, New York, 1969).
6. Faegri, K. & Iversen, J. *Bestimmungsschlüssel für die Nordwestdeutsche Pollenflora* (Fischer, Stuttgart, 1993).
7. Stix, E. *Pollenkalender Stuttgart* (wiss. Verlagsgesellschaft mbH, Stuttgart, 1981).
8. Pohl, F. *Bei. Bot. Cbl.* 56A, 365–470 (1937).
9. Sullivan K. M., Mannucci A., Kimpton, P. A. & Gill, P. *Biotechniques* 15, 636–641 (1993).

Supplementary information is available on Nature's World-Wide Web site (www.nature.com) or as paper copy from the London editorial office of Nature.

Stingray jaws strut their stuff

The cartilaginous skeleton of sharks and rays imposes functional limitations that are not seen in bony fishes. Cartilage is less dense than bone, which helps chondrichthyan (cartilaginous) fishes maintain near neutral buoyancy, but cartilage is also less stiff and strong than bone. Nevertheless, some stingrays routinely use their cartilaginous jaws and pavement-like dentition to crush hard prey, such as snails and mussels¹. We have studied the cownose ray, *Rhinoptera bonasus*, to investigate how cartilaginous jaws can be used to crush hard-shelled prey. The jaws are composed of 'trabecular cartilage', a material that is structurally and functionally convergent with the trabecular bone found in osteichthyan (bony) fishes and tetrapods.

The cownose ray is pelagic, travelling in large schools that periodically descend to beds of sea-grass to feed on benthic invertebrates. The tooth plates used to crush these hard-shelled prey are similar to the teeth of other vertebrates, but the jaws that bear the teeth are entirely cartilaginous. Superficially, the jaws appear to be made of the calcified cartilage that is common to most chondrichthyan fishes, in which a core of hyaline cartilage is surrounded by a layer of prismatic cartilage arranged in tiny mineralized blocks called tesserae^{2,3}. The jaws of the cownose ray exhibit this surface calcification, but cross-sections of the cartilage

show that mineralized struts, composed of columns of tesserae, also run through the centre of the tissue (Fig. 1). The struts are concentrated in the region where the tooth plates crush prey, and are well positioned to distribute the bite force load to the layers of tesserae on the opposite side of the jaw.

In most animals, bones are generally stiffened either by thickening the outer layer of compact bone (cortex) or by increasing the internal trabeculation. The struts, or trabeculae, in the marrow cavity of mammalian long bones follow the lines of stress imposed during normal loading⁴, for example. The jaws of the cownose ray contain analogues of both of these structural modifications.

The analogue to cortical thickening is the deposition of multiple layers of tesserae on the surface of the cartilage (Fig. 1d, red arrows). This character had been thought to exist only in extinct sharks⁵, but the examination of large (> 4 m long) carcharhinid and lamnid sharks revealed as many as seven layers of mineralization at the jaw articulation⁶. Two layers of tesserae cover the entire surface of the jaw in the cownose ray, with some areas having up to six layers.

We were surprised to find that cownose rays also stiffen their jaws with internal trabeculation. Because cartilaginous skeletal elements have no blood supply, they usually exhibit only surface mineralization. Examination of neonatal cownose rays, before

they begin to feed exogenously, shows that trabeculae are already present. This finding, together with the pattern of trabeculation in large individuals, suggests that the trabeculae grow in length by adding tesserae to the ends of the columns, and increase in complexity by branching at the surface and then increasing in length as the jaws grow.

The biochemistry of trabecular cartilage is distinguishable from hyaline cartilage only by its mineral content (25%), which was higher than normal in the matrix, even with the trabeculae removed. The glycosaminoglycan (4.5% of organic material), collagen (27%) and elastin (0%) contents are all within expected ranges.

Biomaterials testing of trabecular cartilage supports the idea that the trabeculae function to stiffen the cartilage. Small blocks (4-mm cubes) of trabecular cartilage were excised and subjected to mechanical tests of stiffness. To remove the mechanical effects of trabeculae, some blocks were treated with EDTA to decalcify the trabeculae, and others were tested with the struts perpendicular to the applied load rather than parallel to it. Both of these groups were an order of magnitude less stiff (10^6 Pa as opposed to 10^7 Pa) than those with intact struts tested parallel to the load.

To our knowledge, this is the first time that organized, mineralized structures serving a distinct biomechanical role have

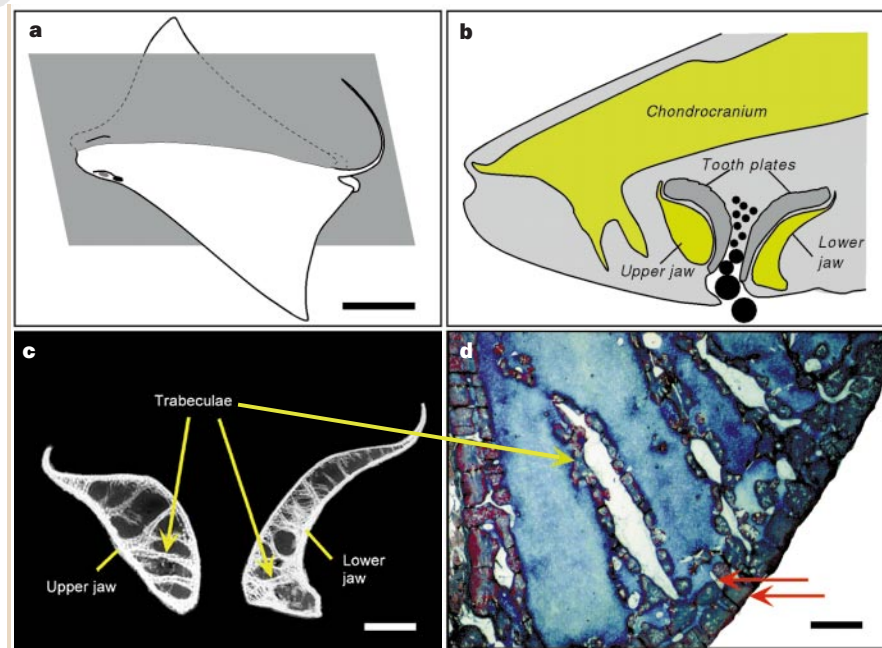


Figure 1 Trabeculae of the cownose stingray, *Rhinoptera bonasus*. **a**, Outline showing the plane through which the head and jaws are sectioned in the other panels. Scale bar, 10 cm. **b**, Sagittal section through the head showing the position of the jaws relative to hard prey items (black circles) that are being crushed. Cartilage is shown in yellow. **c**, Radiograph of a sagittal section through the jaws showing the position of the mineralized struts (trabeculae). The trabeculae are concentrated in the region where the tooth plates crush prey. Scale bar, 1 cm. **d**, Section through the upper jaw after Masson's staining showing that each trabecula, cut obliquely, is a fluid-filled tube. The walls of each trabecular tube are composed of small blocks of mineralized tissue. There is no epithelial tissue lining the tubes and no cellular material appears in the lumen. Most cartilaginous fishes have a single layer of tesserae in the outer prismatic layer of cartilage, but this species has at least two layers (red arrows), and sometimes as many as six. Scale bar, 1 mm.

been found in the core of a mature cartilaginous tissue. Studies of the palaeohistology of vertebrates have assumed that cartilage can mineralize only on the surface. Any new descriptions of fossilized hard tissues, particularly from fishes, must take this discovery into account, and previous descriptions may need to be re-evaluated.

Adam P. Summers*, Thomas J. Koob†, Elizabeth L. Brainerd*

*Organismic and Evolutionary Biology Program, University of Massachusetts at Amherst, Amherst, Massachusetts 01003, USA
e-mail: summers@bio.umass.edu
†Shriners' Hospital for Children, Tampa, Florida 33612, USA

1. Bigelow, H. B. & Schroeder, W. C. *Fishes of the Western North Atlantic* (Sears Foundation for Marine Research, New Haven, 1953).
2. Applegate, S. P. *Sharks, Skates and Rays* (Johns Hopkins Univ. Press, Baltimore, Maryland, 1967).
3. Kemp, N. E. & Westrin, S. K. *J. Morphol.* **160**, 75–102 (1979).
4. Swartz, S. M., Parker, A. & Huo, C. *J. Exp. Biol.* **201**, 573–590 (1998).
5. Schaeffer, B. *Bull. Am. Mus. Nat. Hist.* **169**, 1–120 (1981).
6. Dingerkus, G., Seret, B. & Guilbert, E. *Experientia* **47**, 38–40 (1991).
7. Kirsch, T. & von der Mark, K. *Bone Mineral* **18**, 107–117 (1992).
8. Glimcher, M. J. *Handbook of Physiology: Endocrinology* (American Physiological Society, Washington DC, 1976).

The ubiquitin pathway in Parkinson's disease

Mutations of the α -synuclein gene^{1,2} have been identified in some familial forms of Parkinson's disease, and α -synuclein protein has been shown to accumulate in the brains of patients with the disease³. These findings suggest that Parkinson's disease may be caused by the abnormal aggregation of α -synuclein protein. Here we have identified in a German family with Parkinson's disease a missense mutation in the ubiquitin carboxy-terminal hydrolase L1 (UCH-L1) gene. We show that this mutation, Ile93Met, causes a partial loss of the catalytic activity of this thiol protease, which could lead to aberrations in the proteolytic pathway and aggregation of proteins.

UCH-L1 is one of the most abundant proteins in the brain^{4,5}, comprising up to 2% of total brain protein. Immunoreactivity for this protein is found in Lewy bodies⁶. It belongs to a family of deubiquitinating enzymes, and is thought to cleave polymeric ubiquitin to monomers and to hydrolyse bonds between ubiquitin molecules and small adducts such as glutathione and cellular amines⁷. The abundance of UCH-L1 in human brain, its presence in Lewy bodies, and its involvement in the ubiquitin-dependent proteolytic pathway implicate it in the pathogenesis of Parkinson's disease.

We have sequenced the coding region of the UCH-L1 gene in probands from 72

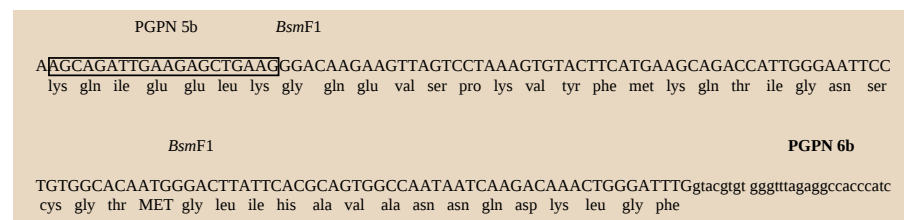


Figure 1 DNA sequence of a portion of exon 4 of the UCH-L1 gene. Boxes indicate polymerase chain reaction (PCR) primer sequences, and arrows indicate restriction sites for the BsmF1 restriction endonuclease.

families with Parkinson's disease, and identified a missense mutation in the fourth exon of the UCH-L1 gene that changes an isoleucine at position 93 to a methionine in one proband of a German pedigree. This Ile93Met change can be examined by a restriction endonuclease assay, because at the nucleotide level the C277G change introduces a new BsmF1 site (Fig. 1). Mutation analysis of the affected brother of the proband showed that he too carries the Ile93Met mutation (Fig. 2).

In both patients, the clinical syndrome is typical for Parkinson's disease. Symptoms began with resting tremor at age 51 for the proband and 49 for her affected brother, and progressed to rigidity, bradykinesia and postural instability. Both individuals showed a beneficial response to L-dopamine replacement therapy. A paternal uncle and the paternal grandmother were also affected, although, with the exception of the two siblings, all other individuals in the pedigree are deceased. The lack of phenotype in the father indicates that the mutation has incomplete penetrance in this family.

We analysed 500 chromosomes from control individuals of different ethnic backgrounds, 204 originating from German backgrounds. None of the 500 control chromosomes examined carried the Ile93Met change. Ile 93 is conserved in UCH-L1, UCH-L3 and the rat and mouse orthologues, as well as in the homologous genes of a yeast and the plant *Arabidopsis thaliana*. Thus, like mutations in α -synuclein, the Ile93Met mutation in the UCH-L1 gene is expected to contribute to the genetic aetiology of only a small number of patients with the familial form of the illness.

The mutant and wild-type proteins were expressed in *Escherichia coli* and assayed using two types of substrate. For the ubiquitin ethyl ester, rate measurements showed that the Ile93Met mutant protein cleaved the substrate (15 μ M) at a rate of 2.41 U mg^{-1} compared with 4.08 U mg^{-1} for the wild-type protein. Because the K_m for this substrate is submicromolar, these rates represent V_{max} values and are consistent with previously determined rates⁸. With the substrate ubiquitin-7-amido-4-methylcoumarin (Ub-AMC)⁹, the V_{max} of the mutant enzyme was 0.20 U mg^{-1} , compared with 0.47 U mg^{-1} for the wild-type enzyme. Mutant and wild-type proteins

exhibited similar K_m values, however, indicating that the mutant protein does not show decreased affinity for the substrate. Similarly, ubiquitin was an equally potent inhibitor of both the mutant and normal enzymes. The enzymatic activity values are consistent with Ile93Met UCH-L1 having a lower catalytic activity than the wild-type protein. Molecular modelling of the mutation suggests alteration in the geometry and fluctuation of the active site.

The roughly 50% reduction in catalytic activity of the Ile93Met protein should be interpreted with caution, however, as the natural substrate for the abundant UCH-L1 protein is not known. The reduced catalytic activity may affect the cleavage and turnover of the unknown substrate(s), leading to aggregation over time of the substrate(s), which can in turn act as a seed for other aggregation-prone abundant proteins. Alternatively, the Ile93Met substitution may render UCH-L1 prone to aggregation with the result that the protein accumulates. Finally, both models — reduced enzymatic activity and enhanced aggregation — may be in play at different stages of the illness.

The finding of mutations in the genes encoding α -synuclein and UCH-L1, and the identification of these proteins in Lewy

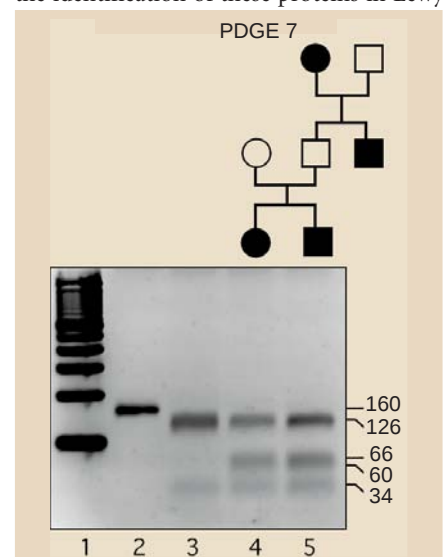


Figure 2 Mutation analysis for the Ile93Met mutation in kindred PDGE7. Lane 1 is a molecular marker; lane 2, undigested PCR product; lane 3, digested PCR product not carrying the mutation; lanes 4 and 5, PCR products from affected individuals carrying the Ile93Met mutation.

bodies in Parkinson's disease, indicate that aberrations in the folding, processing and degradation of proteins lead to neuronal degeneration.

Elisabeth Leroy*, Rebecca Boyer*, Georg Auburger*†, Barbara Leube†, Gudrun Ulm‡, Eva Mezey§, Gyongyi Harta§, Michael J. Brownstein||, Sobhanadditya Jonnalagada¶, Tanya Chernova¶, Anindya Dehejia*, Christian Lavedan*, Thomas Gasser#, Peter J. Steinbach☆, Keith D. Wilkinson¶, Mihael H. Polymeropoulos*

*Genetic Disease Research Branch, NHGRI, NIH, Building 49 Room 4A66, Bethesda, Maryland 20892, USA

†Department of Neurology, University Hospital, PO Box 101007, 40001 Düsseldorf, Germany

‡Paracelsus-Elena-Klinik, 34128 Kassel, Germany

§Basic Neuroscience Program, NINDS, NIH, Building 36 Room 3D06, Bethesda, Maryland 20892, USA

||Section on Genetics, NIMH, NIH, Building 36 Room 3D06, Bethesda, Maryland 20892, USA

¶Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322, USA

#Neurologische Klinik, Klinikum Grosshadern, Ludwig-Maximilians-Universität, München, Germany.

☆Center for Molecular Modeling, CIT, NIH, Building 12A Room 2041, Bethesda, Maryland 20892, USA

1. Polymeropoulos, M. H. *et al. Science* **276**, 2045–2047 (1997).
2. Kruger, R. *et al. Nature Genet.* **18**, 106–108 (1998).
3. Spillantini, M. G. *et al. Nature* **388**, 839–840 (1997).
4. Wilkinson, K. D. *et al. Science* **246**, 670–673 (1989).
5. Wilkinson, K. D., Deshpande, S. & Larsen, C. N. *Biochem. Soc. Trans.* **20**, 631–637 (1992).
6. Lowe, J., McDermott, H., Landon, M., Mayer, R. J. & Wilkinson, K. D. *J. Pathol.* **161**, 153–160 (1990).
7. Larsen, C. N., Krantz, B. A. & Wilkinson, K. D. *Biochemistry* **37**, 3358–3368 (1998).
8. Larsen, C. N., Price, J. S. & Wilkinson, K. D. *Biochemistry* **35**, 6735–6744 (1996).
9. Dang, L. C., Melandri, F. D. & Stein, R. L. *Biochemistry* **37**, 1868–1879 (1998).

Whale ankles and evolutionary relationships

There are two main hypotheses for the relationships of the mammalian order Cetacea (comprising whales, dolphins and porpoises). The first hypothesis, mainly supported by DNA sequence data^{1,2}, is that one of the groups of artiodactyls (for example, the hippopotamids) is the closest extant relative of whales and that Artiodactyla are paraphyletic if Cetacea are excluded from it. The second hypothesis, mainly supported by palaeontological data^{3,4}, identifies mesonychians, a group of extinct archaic ungulates, as the sister group to whales. These two hypotheses are not mutually exclusive, because mesonychians and cetaceans could

be sister groups, and this combined clade (Cete) could be the sister group to a group of artiodactyls.

The morphology of the ankle can be used to evaluate these hypotheses. Ankle specializations are universally used to characterize Artiodactyla, and would provide an excellent test for the inclusion of whales in that order. Unfortunately, the few cetacean ankle bones known are too incomplete or too reduced to allow meaningful comparison with other mammals.

We have recently recovered fragmentary Eocene astragali (ankle bones) from pakicetid and ambulocetid cetaceans⁵ in Pakistan. We identified them as cetaceans because the deeply grooved trochlea resembles the partial astragalus of the holotype of *Ambulocetus natans*⁵, and the large size of the astragali matches only a few mammals known from the associated freshwater and marine faunas, in particular perissodactyls, anthracobunids and sirenians. Astragali for known representatives or relatives of these mammals do not match the morphology of the new bones. The ambulocetid astragalus was found in marine sediments.

Three articular facets of artiodactyl ankles are highly specialized and are important for the relationship of whales. First, the astragalar head of artiodactyls is trochleated, meaning that it is wide, gently concave mediolaterally, and strongly convex dorso-plantarily, with the axis of this convexity perpendicular to the median plane. Second, the sustentacular facet is rectangular and covers the entire posterior aspect of the astragalus. Finally, the ectal (posterior calcaneo-astragalar) facet is reduced and placed on the lateral side of the bone. The combination of these features is found in all artiodactyls but not in any other mammal.

The cetacean astragalar head (Fig. 1) is

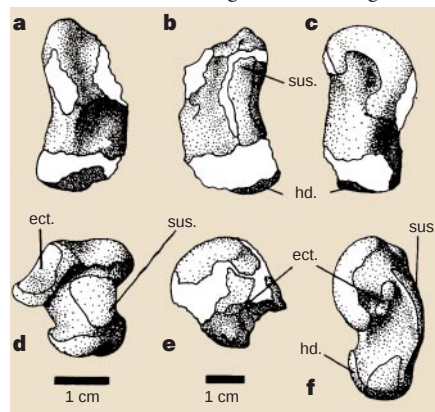


Figure 1 Astragali of: **a–c**, pakicetid cetacean shown in dorsal (**a**), plantar (**b**) and medial (**c**) view (H-GSP 97227, Locality 300); **d**, mesonychian (plantar view, *Disaccus europaeus*, MNHN Br 21 L); **e**, ambulocetid cetacean (lateral view, H-GSP 97113, Locality 9205, distal part missing); and **f**, artiodactyl (lateral view, *Sus scrofa*). Shown are ectal facet (ect.), sustentacular facet (sus.) and astragalar head (hd.). Left scale bar is for **a–d**, right scale bar is for **e–f**.

wide and nearly flat both mediolaterally and dorso-plantarily. This is unlike the condyle of mesonychians, but is also unlike the convex trochleated head of artiodactyls. This important feature, often cited as the main defining character of artiodactyls⁶, is inconsistent with the hypothesis that cetaceans should be included in the artiodactyls.

The cetacean sustentacular facet resembles that of artiodactyls in being long, but unlike that of artiodactyls it is narrow. In primitive mammals, including mesonychians, the sustentacular facet is short and rounded. The cetacean ectal facet is strongly reduced and placed laterally as in artiodactyls, not plantarily as in mesonychians and other primitive mammals. This position of the ectal facet is highly derived⁷ and unique, occurring only in artiodactyls and these Eocene cetaceans. These features argue against close phylogenetic ties between cetaceans and mesonychians.

Our new ankle data do not unambiguously support either of the predominant hypotheses of cetacean relationships. Inclusion of Cetacea in Artiodactyla to the exclusion of mesonychians is consistent with the position of the ectal facet and the shape of the sustentacular facet. But the absence of a trochleated astragalar head argues against the inclusion of Cetacea in Artiodactyla, unless the flat head of the cetacean is interpreted as a secondary aquatic adaptation. Inclusion of Cetacea in Artiodactyla is also inconsistent with the derived similarities of the dentition and basicranium of cetaceans and mesonychians⁸. Sister-group relations between mesonychians and cetaceans are inconsistent with the derived similarities in the sustentacular and ectal facets between artiodactyls and cetaceans, both characters with little or no homoplasy in mammals. But, in any case, extensive convergence or reversals must have occurred in the dentition, basicranium and/or tarsus.

J. G. M. Thewissen

Department of Anatomy, Northeastern Ohio Universities, College of Medicine, Rootstown, Ohio 44272, USA

S. I. Madar

Department of Biology, Hiram College, Hiram, Ohio 44234, USA

S. T. Hussain

Department of Anatomy, Howard University, College of Medicine, Washington DC, 20059, USA

1. Gatesy, J., Hayashi, C., Cronin, M. & Arctander, P. *Mol. Biol. Evol.* **13**, 954–963 (1996).
2. Shimamura, M. *et al. Nature* **388**, 666–670 (1997).
3. Van Valen, L. *Am. Mus. Nat. Hist. Bull.* **132**, 1–126 (1966).
4. Thewissen, J. G. M. *J. Mamm. Evol.* **2**, 157–184 (1994).
5. Thewissen, J. G. M., Madar, S. I. & Hussain, S. T. *Cour. Forsch. Inst. Senck.* **191**, 1–86 (1996).
6. Gentry, A. W. & Hooker, J. J. in *The Phylogeny and Classification of the Tetrapods Vol. 2. Mammals* (ed. Benton, M. J.) 235–272 (Clarendon, Oxford, 1988).
7. Schaeffer, B. *Am. Mus. Novit.* **1356**, 1–24 (1947).
8. Geisler, J. H. & Luo, Z. in *The Emergence of Whales* (ed. Thewissen, J. G. M.) 163–212 (Plenum, New York, 1998).

To see a world in a grain of sand

Destiny or Chance: Our Solar System and Its Place in the Cosmos

by Stuart Ross Taylor
Cambridge University Press: 1998. 256pp.
\$24.95, £17.95

Bernard Lovell

Two hundred years ago Laplace proposed that the Earth and planets evolved from a nebula of gas and dust surrounding the Sun. This remains the basis of nearly all contemporary theories, although the processes through which the gas and dust of the rotating solar nebula eventually became the Solar System as we see it today are still in dispute.

Professor Stuart Taylor of the Australian National University is a planetary scientist well qualified to write on this subject, and he has produced a clear account of the complex occurrences over the past 4 to 5 billion years that have led to the formation of the Earth, Moon and planets. Any contemporary explanation has to take account of the remarkable discoveries of the past few decades, during which samples of lunar rock have been returned to Earth, and space probes have transmitted back hitherto unknown details of the planets and their environments.

An overriding impression from Taylor's account is that major collisions have been an essential factor in producing the present system. It is no longer possible to believe that the planets were formed simply by a steady accretion of the dust and gas of the primordial nebula. The diversity of planets and the heavily cratered surfaces of the inner planets are incompatible with this idea. The modern concept is that the gas and dust of the solar nebula accreted into solid bodies of a few metres or kilometres in size, perhaps within a million years of the formation of the nebula. Over the next 50 million years, the dominant influence in the formation of the planets was the repeated collisions of these rocky bodies. Most of the gas in the nebula would have been swept away within a few million years, but by that time Jupiter and the outer planets were massive enough to attract sufficient of the gas to form the "gas giants" as we see them today.

The continued massive collisions are believed to have been responsible for the strange difference in the tilts and speeds of rotation of the planets. For example, Uranus and Neptune are similar, with days of 17 and 16 hours, respectively. The axis of rotation of Neptune is tilted to the common plane of the Solar System by nearly 30 degrees, but Uranus is lying on its side. At some stage in the evolutionary process, a body of about the size of the Earth must have collided with



Jupiter and its Galilean satellites, Io, Europa, Ganymede and Callisto, photographed by Voyager One.

Uranus to create this odd situation.

At least 60 satellites in orbit around the planets are now known. The giant outer planets — Jupiter, Saturn, Uranus and Neptune — have 57 of them. These have been formed either by capture of the bodies left over after the major accretion and collisional processes, or by accretion in the gaseous disks thrown out from the planets by collisions. Of the inner planets, Mars probably captured its two satellites, Phobos and Deimos. The formation of the Earth's satellite — the Moon — has been a topic of great dispute. Professor Taylor gives a clear description of the present-day opinion that the Moon was formed in an oblique collision with the Earth over 4 billion years ago by a body about the size of Mars. Computer mod-

els show that a few hours after the collision the debris of the impactor was spread out in space, but soon clumped together by gravitational attraction, and after about 24 hours formed the body in orbit around the Earth — our Moon. A small amount of the Earth's mantle ended up as part of the Moon. The heat generated by the impact would have melted the impactor and the Earth's mantle. The analysis of the lunar material returned to Earth by the Apollo astronauts when compared with the constitution of the Earth's mantle seems compatible with this theory.

Taylor's account reveals that Jupiter has been a critical influence in preventing the formation of a planet in the region between Mars and Jupiter. In this region, the gravitational effect of the planet either collected or tossed into the Sun much of the rocky mater-

ial, so that an accretion process became impossible, and in that region we now have the asteroid belt. Some 10,000 asteroids have been discovered, and one thousand of these, of more than one kilometre in size, cross the orbit of the Earth. Professor Taylor's estimate of the chance of another celestial collision, similar to that which led to the extinction of the dinosaurs, places the uniqueness of our existence in perspective. So, too, does his estimate that we would be unlikely to exist but for the fortunate circumstances that led to the formation of the planet Jupiter, which, because of its gravitational effect, has saved us from frequent bombardment by comets.

In the last section of this book, Professor Taylor reveals the reason for his title. He believes that his account shows that only chance events have led to the evolution of the Solar System, of the existence of the Earth and of life. He believes that the sequence of these chance encounters makes life on Earth unique in the Universe and that searches for life elsewhere are futile. Many readers may feel that this last section, with its anti-religious undertones, is an unnecessary and unworthy conclusion to an otherwise easily readable account of the formation of the Solar System. □

Bernard Lovell is at the Nuffield Radio Astronomy Laboratories, University of Manchester, Jodrell Bank, Macclesfield, Cheshire SK11 9DN, UK.

Fat and fiction: time to weigh the benefits

The Fats of Life

by Caroline M. Pond
Cambridge University Press: 1998. 328pp.
£35, \$59.95 (hbk); £12.95, \$19.95 (pbk)

John Prins

Is fat friend or foe? This question has vexed people for an age, but in recent times has been somewhat forgotten, with fat firmly ensconced as villain. In *The Fats of Life*, Caroline Pond aims to redress the balance, championing the cause of fat and its many intimate associates, including lipids, blubber and fat-soluble vitamins.

Pond gathers her evidence from many sources — most prominently her own field of comparative biology, but also archaeology, plant biology, historic writings and even mythology. This can make for fascinating reading, but is occasionally somewhat repetitive. Another problem is that, for much discussion, argument and presented 'fact', no references are given. Indeed, in much of the book, the anecdotes are better referenced than the science — a feature that will irritate the scientific but may not trouble the lay reader, as it is the anecdotes that most beckon further reading.

The book's stated aim is to "fill the gap



Edmond Boubli's "Ronde de Nuit" line on show in Paris.

between unscientific comments about the hazards and benefits of high-fat or low-fat diets and weight control found in magazines, and technical and medical reports about lipid biochemistry and obesity". Thus readers would hope to learn about the role of fat in their own bodies, its likely advantages and disadvantages, and perhaps gain some insight into how their weight may be modified if desired.

Unfortunately, while much of this information can be found in the book, it is rather scattered about. The reader comes away with no clear impression of whether overweight or obesity is truly bad, or if it may even be good. The energy balance equation is discussed or alluded to in various chapters, but its overriding importance is never made clear. For example, in weight control, dietary

fat content only really makes a difference if there is a positive energy balance, meaning that you are consuming more calories than you are using up; but, while this crucial point is well made, it is also well hidden. Furthermore, Pond's arguments on the value of fat often draw on the consequences of extreme low-body-fat situations such as starvation and anorexia. This is no different from extrapolating backwards from studies of individuals with morbid obesity — an extrapolation about which Pond is critical. Similarly, the book concentrates on the metabolic aspects of fat biology and consequences of obesity, as opposed to those that are less (scientifically) trendy, such as psychiatric and orthopaedic problems, and increased rates of malignancy. No one would dispute that, in terms of cardiovascular disease or non insulin-dependent diabetes, it is better to have lipids inside, rather than outside, fat cells. But as many individuals, as well as any rheumatologist, orthopaedic surgeon, plastic surgeon or psychiatrist, will testify, excess lipid stored within fat cells may also have adverse consequences, including osteoarthritis, low self-esteem or depression, and discrimination.

The Fats of Life does not present an altogether balanced view, but there, in fact, lies part of its attraction. Its major contribution may well be to help swing the pendulum back a little, widening the recognition that in order to be healthy, one does not need to be anorexic or nearly so. □

John Prins is in the Departments of Medicine and Clinical Biochemistry, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QR, UK.

Complementing for fishes

The Diversity of Fishes

by G.S. Helfman, B.B. Collette, D.E. Facey
Blackwell Science: 1997. 528pp. £55, \$69.95
Hans Fricke

A fish is not so much a taxonomic term as a convenient description for aquatic organ-

New in paperback

What Emotions Really Are

by Paul E. Griffiths
University of Chicago Press: 1998. 286pp. \$17, £13.50
"It is difficult to do justice to Griffiths in a short review. His analysis of the concept of emotion and his proposal for the future direction of the field is the most compelling and best argued I have read. *What Emotions Really Are* makes a strong claim to be one of the best books to have emerged on the subject of human emotion." Ray Dolan, *Nature* 391, 35–36 (1998).

Mirrors in Mind

by Richard Gregory
Penguin: 1998. 302pp. £12.99
Reviewed by Simon Altmann in *Nature* 385, 785 (1997).

Huxley

by Adrian Desmond
Penguin: 1998. 820pp. £10.99
Part 1 reviewed by W.F. Bynum in *Nature* 372, 300 (1994) and part 2 by Crispin Tickell, *Nature* 386, 349 (1997).



Plenty more fish in the sea: clockwise from top left, lionfish, boxfish, wolf-fish and tuskfish

isms, loosely confined to the vertebrate. In this monumental work, we learn of 49,100 such vertebrates, and that fishes, with more than 50 per cent of the total, do well within the vertebrate world. They are conquerors of the aquatic realm, living almost anywhere, and breaking all sorts of records among the backboneed. They thrive at elevations of 5,200 m in the Himalayas, and at the mephistophelian depths of 8,000 m in the ocean; and they can live in temperatures at the human pain threshold of almost 50 °C, and, thanks to antifreeze in the blood, down to below zero in Antarctic waters. Amphibians, reptiles, birds and mammals cannot compete.

Fish offer an amazing variety of sex lives. They are promiscuous or polygamous, mate for seconds, or for their entire lives; their sex can be fixed, or they may change it. Both sexes can be simultaneously possessed by one individual, or they can alternate from one to the other. Last but not least, one species even omits the male sex and consists of females only, their eggs activated by the sperm of foreign species. Fishes are the kings of the vertebrates. This book was written by those who admire them, and is directed at those with an interest in them. The authors want to turn this interest into fascination, and they succeed.

In the "shameless self-promoting department" of their website, the authors advertise their work as "far the most complete and detailed ichthyology textbook available". *The*

Diversity of Fishes is, indeed, a massive enterprise. The book's 25 readable chapters explore the diversity of fish in terms of anatomy, taxonomy, phylogeny, physiology, ecology and behaviour, even science history, and also, importantly, their future. There are endless figures, tables and boxes, the last taking up such snappy matters as "Just what is a bichir?", "Social transmission of cultural traditions in fishes" and "The snail darter and the politics of endangerment". "We have gone to considerable lengths to get our facts straight or to admit where uncertainties lie", the authors claim, and their honesty extends to a website with updates about misspellings, reversed photographs and art works, faulty figures, and incomplete references.

But having developed in us a tremendous regard for these cold-blooded (except a few of them) "insects" of the aquatic realm, the authors have a surprise for us at the very end. The massive reference section carries 1,800 citations, but not one of them in a language other than English. Are all original publications in Dutch, German, French, Japanese and Russian (most of them with English summaries) below standard? For example, where is the Nobel laureate Karl von Frisch's classic work on hearing in fish, with the refreshing title *Ein Fisch der kommt wenn man pfeift* (A fish which comes if one whistles)? A box for von Frisch would have been nice. There are many other works of this calibre in the international non-English literature. □

Hans Fricke is at the Max Planck Institute for Animal Behaviour, 82319 Seewiesen, Germany

Impurities in the historical stream

The Bends: Compressed Air in the History of Science, Diving, and Engineering

by John L. Phillips

Yale University Press, New Haven and London, 1998. 217pp. plus Notes, Bibliography and Index, \$30.00

James Vorosmarti

John Phillips has packed a great deal of information from a wide range of sources into this history of compressed air. For those who have worked and studied in this field none of it is new — contrary to the claims of the dust-jacket, there is no evidence that the author drew on previously unpublished letters, diaries and notes, and, ironically, the book is not very informative about the bends itself. But in general, *The Bends* is well done.

The author has for the most part correctly identified the major advances in the use of compressed air in diving and caisson work and those who made them. The chapters dealing with Charles-Jean Triger, James Eads and the first bridge across the Mississippi, the physician Alphonse Jaminet (who contracted the bends), and the Brooklyn Bridge are all well written and provide interesting reading to those who are not familiar with the pioneering engineering of this area of civil engineering.

What is irritating and detracts from the volume are the many examples of careless writing or editing. In one instance, Phillips states that compressed air was invented in the 1840s, when there is evidence in the text that compressed air was known well before this. Thomas Smeaton did invent the first practical air compressor, but he did not unite compressors with steam power. Phillips ignores F. Hoppe, the first scientist to show that on decompression free gas could be found in the heart and the great veins leading to it, and, in 1857, rightly attributed death in caisson workers to this liberation of gas from the blood and tissues. This was before Paul Bert's similar demonstration, and his discovery that the bubbles were mainly nitrogen.

Phillips does not include Bert's important conclusions that oxygen and recompression would be useful in the treatment of decompression sickness (DCS). He repeats the myth that Augustus Siebe invented the diving helmet in 1819 (invented by Charles and James Deane in 1828), although Siebe did produce the first closed diving dress in 1839, which was the model for deep-sea diving dress into the 1970s. He first uses the term SCUBA for the Denayrouse-Royquayrol apparatus of 1865, while missing the fact that

Charles Condert (discussed in the text) demonstrated this type of apparatus in 1832. There are many other examples.

The clarity of Phillips's thinking is also occasionally called into question. For example, in a discussion of how the present atmosphere evolved, he says: "Nitrogen levels gradually rose during this time but, because nitrogen was non-toxic, it tended to higher and higher partial pressures"; and "Only after the physiology of the lungs at sea-level and the laws governing the atmosphere itself could be understood, could the lungs function successfully using compressed air below

sea level." He states that the history of DCS can serve as a model for understanding other occupational or engineering-induced diseases that may arise in the future, but there is no evidence for this.

A more fundamental defect is the author's obvious lack of training or experience in the field and his poor grasp of basic knowledge of the pathophysiology of DCS or air embolism; when he attempts to interpret what he has gleaned the picture is very muddled. In the last chapter, he gives 10 major statements regarding DCS: these are very naive and in some cases not accepted by

those in the field.

The Bends will be interesting to those who are not familiar with the topic and want a short introduction to the history of diving and caisson work. It is not recommended for anyone who has a knowledge in the area, who would find the mistakes, omissions and misinterpretations irritating. The list of references is reasonably complete except for the glaring omission of R.H. Davis's classic *Deep Diving and Submarine Operations* which contains most of the historical material in this text and would be useful to those who want to learn more. □

In retrospect chosen by Roger Short

An Essay on the Principle of Population

by Thomas Robert Malthus
(1798)

Edited with an Introduction by Antony Flew
(Penguin, 1982)

The Revd Thomas Robert Malthus is surely one of the most maligned and misunderstood of men. It is exactly 200 years since the publication of his first, anonymous edition of *An Essay on the Principle of Population, As It Affects the Future Improvement of Society*. The stimulus for writing this polemic was his concern about the unwarranted euphoria of his colleagues, who in the aftermath of the French Revolution saw mankind progressing ever upwards to a world of universal abundance, peace and prosperity, where all would be equal in health, wealth and happiness. He wished to de-bunk this utopian fantasy, and used his numeracy (he had graduated from Cambridge with the equivalent of a First in Mathematics) to point out a simple truth:

The power of population is indefinitely greater than the power in the earth to produce subsistence for man. Population, when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will shew the immensity of the first power in comparison to the second. By that law of our nature which makes food necessary to the life of man, the effects of these two unequal powers must be kept equal. This implies a strong and constantly operating check on population from the difficulty of subsistence. This difficulty must fall somewhere and must necessarily be severely felt be a large portion of mankind.

This "Dismal Theorem", published when Malthus was only 32, was viciously attacked by Karl Marx, Frederick Engels and other like-minded social reformers of the day, who could not accept the idea that the poor would always be with us. But at the same time it was welcomed by Charles Darwin and Alfred Russell Wallace, as it showed them how survival of the fittest might ultimately result in the evolution of new species.

Alas, Malthus's pessimistic forebodings



about the difficulty of subsistence have been verified by subsequent events. Today, according to the World Bank, over one billion people are living in absolute poverty, and 600 million are on the borders of starvation. But contrary to Malthus's predictions, there has been an exponential growth in population. He was writing at a time when the world's population was a mere one billion; he would never have guessed that today it would have increased to around six billion, and is projected to reach 9–10 billion by the year 2050. Whatever happened to Malthus's "strong and constantly operating check on population"?

Summarizing his views on "The Principle of Population" in 1830, Malthus concluded that the checks could be of two kinds. They were either voluntary and "preventive", through late marriage and sexual restraint (Malthus was an exponent of both), or they were "positive", as exemplified by "promiscuous concubinage" (contraception), "improper arts to conceal the consequences of irregular connections" (abortion), and "wars, infanticide, plague and

famine" — in short, a catalogue of misery and vice. Malthus's condemnation of contraception within marriage was only reversed by the Anglican Church in 1930.

Malthus, who was still a bachelor in 1798, was understandably unaware of the fact that lactational amenorrhoea, the inhibitory effects of suckling on ovulation, is nature's most important check on human fertility. The erosion of traditional breastfeeding practices by urbanization, and by the premature introduction of animal milk or infant milk, has been one of the major factors that has stimulated human fertility; this, coupled with the conquest of infectious diseases in infancy and childhood, has fuelled the exponential growth of the human population.

I have often wondered why Malthus chose to publish the first edition of his essay anonymously. Was it all just too embarrassing for the shy young bachelor don of Jesus College, Cambridge, who had only recently taken Holy Orders? Was his discussion of the essential nature of "the passion between the sexes" too strong meat for his celibate colleagues at High Table? How were Malthus's views on population received by the young gentlemen of the East India Company, attending its college at Haileybury where Malthus held the Professorship in Political Economy (the first in Britain) from 1805 until his death in 1834? Did it ever dawn on them that increasing human numbers would eventually make India ungovernable by a small colonial power such as Britain?

At the end of the day, we should remember Malthus for posing the unanswerable question to all the world's political economists — how far can we afford to go to alleviate poverty? His "Dismal Theorem" will haunt us in the coming century, as the gulf between developed and developing nations continues to widen, and the poor struggle to scrape a living from the surface of our increasingly denuded planet. The help we are prepared to give to those less fortunate than ourselves will be the ultimate measure of our humanity.

Roger Short is in the Department of Perinatal Medicine, Royal Women's Hospital, University of Melbourne, Victoria 3053, Australia

Electron–positron jets associated with the quasar 3C279

J. F. C. Wardle, D. C. Homan, R. Ojha & D. H. Roberts

Department of Physics, Brandeis University, Waltham, Massachusetts 02254, USA

A long-standing question in extragalactic astrophysics is the composition of the relativistic jets of plasma that stream from the nuclei of quasars and active galaxies—do they consist of a ‘normal’ (electron–proton) plasma, or a ‘pair’ (electron–positron) plasma? Distinguishing between these possibilities is crucial for understanding the physical processes occurring close to the putative supermassive black holes that are believed responsible for the jets. Here we report the detection of circularly polarized radio emission from the jets of the archetypal quasar 3C279. The circular polarization is produced by Faraday conversion, which requires the energy distribution of the radiating particles to extend to very low energies, indicating that electron–positron pairs are an important component of the jet plasma. Similar detections in three other radio sources suggest that, in general, extragalactic radio jets are composed mainly of an electron–positron plasma.

The most luminous objects in the Universe are distant quasars and active galactic nuclei. A few per cent of these emit strongly at radio wavelengths, and are powered by twin jets of plasma streaming at close to the speed of light from the nucleus of the underlying galaxy¹. The central ‘engine’ in the nucleus is widely believed to be a supermassive black hole of some 10^8 – 10^9 solar masses². At radio wavelengths the jets are visible from parsec to kiloparsec scales, emitting synchrotron radiation from ultra-relativistic electrons gyrating in a magnetic field. The bulk kinetic energy carried by the jets powers the extended lobes of radio emission far outside the parent galaxy, while synchrotron and inverse Compton emission from the base of the jets generates radiation at optical, X-ray and even γ -ray wavelengths^{3,4}. The composition of the jet plasma has been an unresolved issue ever since the discovery of the jets. The two main candidates are a ‘normal’ plasma consisting of protons and relativistic electrons (an e–p jet), and a ‘pair plasma’ consisting only of relativistic electrons and positrons (an e^+e^- jet).

Electrons and positrons emit synchrotron radiation of identical spectrum and linear polarization, so it is difficult to tell whether the radiating particles are mainly electrons or a mixture of electrons and positrons. Two lines of argument shed light on the problem. The first comes from matching the total energy flux carried by the jet to the total dissipation observed in electromagnetic radiation—the energy stored in the extended radio lobes in magnetic fields and relativistic particles—and the mechanical work done against the surrounding medium^{5,6}. The energy distribution of the radiating particles is assumed to be a power law, $n(\gamma) \propto \gamma^{-(2\alpha+1)}$ for $\gamma_{\min} < \gamma < \gamma_{\max}$, where n is the density, $\gamma = (1 - v^2/c^2)^{-1/2}$ is the Lorentz factor of a particle of velocity v , α is the observed spectral index of the synchrotron radiation, and c is the velocity of light. For an e–p jet, the energy distribution must cut off at $\gamma_{\min} \approx 100$, or the jet may carry several orders of magnitude more energy than is seen to be dissipated. For an e^+e^- jet, the energy distribution must extend to low energies, $\gamma_{\min} \approx 1$, in order that the jet carries sufficient energy. The second line of argument is based on the strong fractional linear polarization often observed in the jets on parsec scales^{7,8}. This strongly limits the amount of internal Faraday rotation (due mainly to the lowest-energy electrons in the jet). Interestingly, the result is similar. For an e–p jet, $\gamma_{\min} \geq 100$, while for an e^+e^- jet there is no Faraday rotation (since electrons and positrons gyrate in opposite directions) and hence there is no constraint on $\gamma_{\min}$.

Thus a value of $\gamma_{\min} \ll 100$ would be strong evidence in favour of an e^+e^- jet, whereas $\gamma_{\min} \geq 100$ would require an e–p jet. The lowest-energy radiating particles are not directly observable,

because their radiation is strongly self-absorbed, though the signature of a low energy cut-off may be visible in inverse Compton scattered radiation at X-ray and γ -ray wavelengths. Here we present a new approach to the question of jet composition by imaging the circularly polarized component of the synchrotron emission. We have detected circular polarization from the archetypal violently variable quasar 3C279, at several epochs, using the Very Long Baseline Array (VLBA) at 15 GHz. We argue that the circular polarization is produced by Faraday conversion^{9,10} of linear to circular polarization in the jet plasma, and requires the energy distribution of the radiating particles to extend to $\gamma_{\min} \ll 100$, implying an electron–positron jet.

The quasar 3C279, at a redshift of $z = 0.538$, is one of the most luminous objects in the sky from radio to γ -ray wavelengths, and is violently variable across the electromagnetic spectrum^{12,13}. At radio wavelengths, images obtained^{14–16} using very-long-baseline interferometry (VLBI) show a bright unresolved core and a one-sided jet which extends to kiloparsec scales¹⁷, ending in diffuse emission. It was the first source in which superluminal motions in the jet were observed¹⁸, and subsequent observations have measured apparent velocities from 4 to 15 times the speed of light (we assume $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.05$). Superluminal motion indicates both bulk relativistic speeds in the jet, and that it makes a small angle to the line of sight. It follows that the intensity of the observed radiation from the jet at all wavelengths is strongly boosted by the relativistic Doppler effect. Its high flux density in all wavebands makes 3C279 one of the most extensively observed of all quasars, and a good system in which to examine the physics of extragalactic jets.

The observations

The observations were part of a programme to monitor the structure and polarization of 12 rapidly variable compact radio sources, using the VLBA at 22 GHz (1.3 cm wavelength) and 15 GHz (2.0 cm wavelength). Observations were made at intervals of two months, and the consistency between images made at different epochs is an important check on our results. Here we present the July 1996 15-GHz observations of 3C279: our techniques, and some of the tests and simulations we have performed, are given in Methods. We will present elsewhere the observations of 3C279 at other epochs, and the observations of the sources 3C84, 3C273 and PKS0528+134, in which we also detected circular polarization.

Images of 3C279 in total intensity, linear polarization and circular polarization are shown in Fig. 1. At a redshift of 0.538, 1 mas corresponds to a linear scale of 5.9 pc.

Figure 1c shows the circularly polarized image. The circular-polarization-emitting region is in the core, slightly displaced to the west, and has an integrated flux density of $V = +89 (\pm 10)$ mJy. Compared to the noise level on the rest of the image, this is a 40σ detection, confirmed by several independent calibration techniques and diagnostic tests (see Methods), and it is seen at every epoch for which we have good data.

The measured component properties are listed in Table 1. The two components in the core region, CE and CW, probably represent regions of a quasi-continuous inhomogeneous jet, as described by Blandford and Königl¹⁹. We interpret CE as the optically thick base of the jet, and CW as being close to the optical surface, where the jet becomes transparent, possibly enhanced by a newly emerging component. Assuming that the circularly polarized signal is associated with component CW (the higher opacity of CE strongly depresses any circular polarization), its fractional circular polarization is $m_C = +1.0 (\pm 0.2)\%$. At 22 GHz, the data are significantly

noisier due to tropospheric fluctuations and beam squint problems. The “zero V self-calibration” and “phase-only mapping” techniques (see Methods) both show no circularly polarized signal above 50 mJy, corresponding to a fractional polarization of CW at 22 GHz of $m_C \leq 0.5\%$.

The absolute polarization position angle calibration was determined by setting the position angles for component K1 (see Fig. 1) equal to 67° (parallel to the jet direction) at both 15 and 22 GHz. This is consistent with the calibration determined by comparing the VLBA observations of OJ287 with nearly simultaneous VLA observations. The polarization position angles of CW imply that there is $8^\circ (\pm 3^\circ)$ of Faraday rotation between 15 and 22 GHz (also observed at other epochs). This corresponds to a Faraday depth at 15 GHz of $\tau_F = -1.1 \pm 0.4$ (observed rotation measure, $-700 (\pm 260)$ rad m^{-2}). This is an important constraint on the line-of-sight component of the magnetic field, and is crucial for the interpretation of the circular polarization.

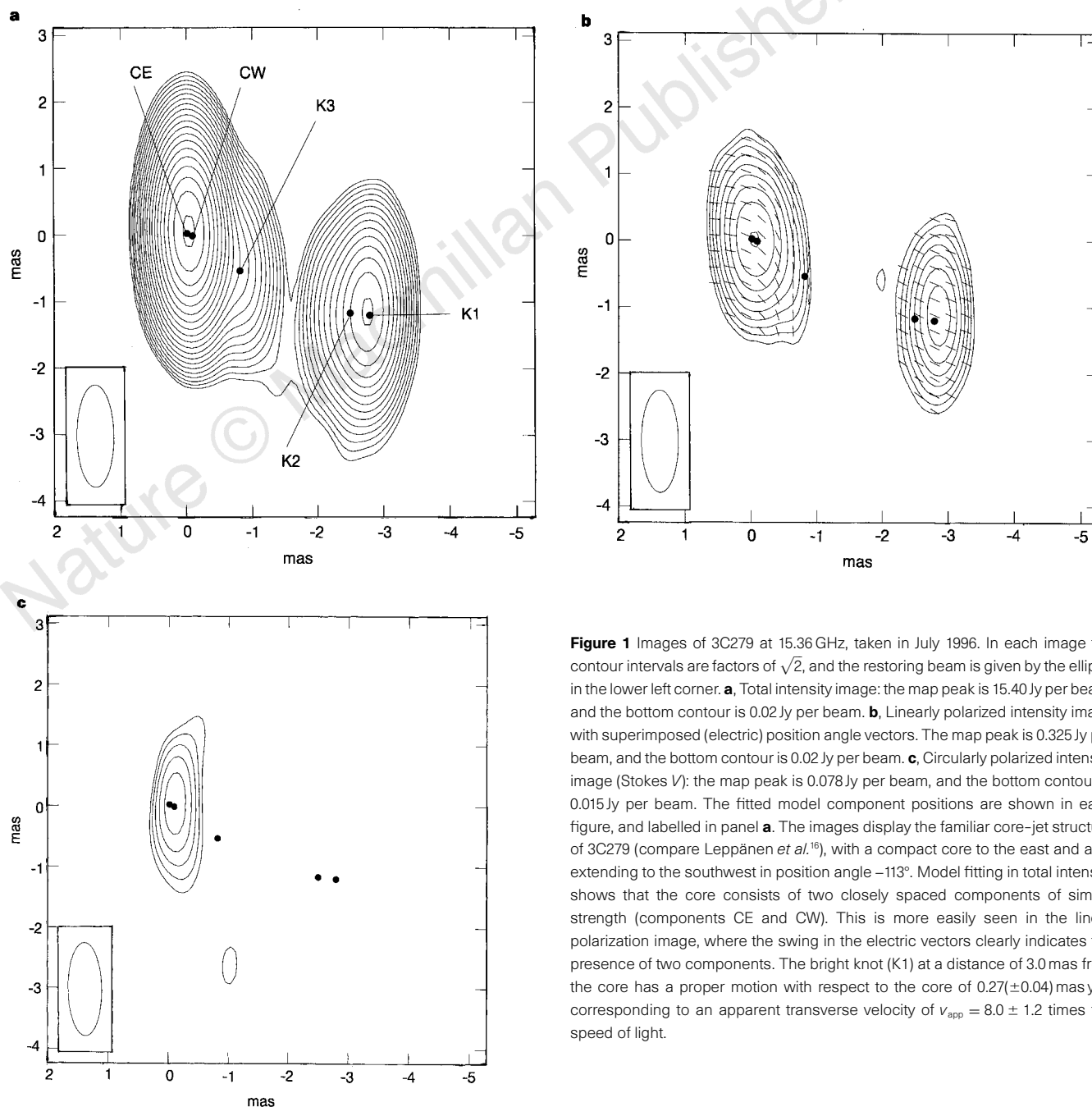


Figure 1 Images of 3C279 at 15.36 GHz, taken in July 1996. In each image the contour intervals are factors of $\sqrt{2}$, and the restoring beam is given by the ellipse in the lower left corner. **a**, Total intensity image: the map peak is 15.40 Jy per beam, and the bottom contour is 0.02 Jy per beam. **b**, Linearly polarized intensity image with superimposed (electric) position angle vectors. The map peak is 0.325 Jy per beam, and the bottom contour is 0.02 Jy per beam. **c**, Circularly polarized intensity image (Stokes V): the map peak is 0.078 Jy per beam, and the bottom contour is 0.015 Jy per beam. The fitted model component positions are shown in each figure, and labelled in panel **a**. The images display the familiar core–jet structure of 3C279 (compare Leppänen *et al.*¹⁶), with a compact core to the east and a jet extending to the southwest in position angle -113° . Model fitting in total intensity shows that the core consists of two closely spaced components of similar strength (components CE and CW). This is more easily seen in the linear polarization image, where the swing in the electric vectors clearly indicates the presence of two components. The bright knot (K1) at a distance of 3.0 mas from the core has a proper motion with respect to the core of $0.27 (\pm 0.04)$ mas yr^{-1} , corresponding to an apparent transverse velocity of $v_{app} = 8.0 \pm 1.2$ times the speed of light.

Interpretation

For simplicity, we shall treat CW as a single homogeneous region in the jet, and assume that the optically thin spectral index of CW has the canonical value $\alpha = 0.5$. Our results are not qualitatively different if another value is used. Both the spectrum and the linear polarizations for CW listed in Table 1 are best fitted by an optical depth at 15 GHz of about $\tau \approx 0.7$ over a range of magnetic-field models, and we shall fix the optical depth at this value. Then, given the observed brightness temperature, we can derive the magnetic field strength using standard formulae²⁰. Model fitting gives the size of CW as $(0.34 \text{ mas}) \times (< 0.1 \text{ mas})$, elongated along the jet, and we parametrize the transverse size of CW as $0.1\eta \text{ mas}$. If the opening angle of the jet is taken from components K1 and K2 (see Fig. 1), then the transverse dimension is 0.01 mas (that is, $\eta = 0.1$), and we consider this to be a lower limit on η . We find that $B_{\perp} = 5.8\eta^2\delta \text{ mG}$, where $B_{\perp}$ is the root-mean-square value of the magnetic field in the plane of the sky, and $\delta = [\gamma_{\text{jet}}(1 - \beta_{\text{jet}}\cos\theta)]^{-1}$ is the Doppler factor of the radiating fluid (the jet speed is $\beta_{\text{jet}}c$, $\gamma_{\text{jet}} = (1 - \beta_{\text{jet}}^2)^{-1/2}$ is its Lorentz factor, and θ is the inclination of the jet to the line of sight). A synchrotron self-Compton calculation²¹ gives $\delta \approx 20$, while the observed proper motion of K1 requires $\delta > 8$. These two estimates bracket a plausible range for δ . The average Lorentz factor of electrons radiating at an observed frequency of 15.36 GHz is $\gamma = (1.2 \times 10^3)/\eta\delta$.

To interpret the circular polarization, we must also assume a structure for the magnetic field, and this is strongly constrained by the observed linear polarization. A plausible family of field structures is given by the “shock in jet” models^{22–24}, in which a shock compresses the jet fluid, partially ordering the magnetic field transverse to the jet direction. The linear polarization of the jet (Fig. 1b) shows that the electric vectors are generally aligned with the jet direction, indicating a predominantly transverse magnetic field. This is consistent with a shock model, but our results do not depend on this. It is simply a convenient way to describe a general magnetic-field structure that contains both a uniform component and a partially ordered tangled component.

The production and transfer of circular polarized radiation (in a uniform magnetic field) has been analysed in detail by Jones and O’Dell⁹, who give the general solution for the four Stokes parameters at arbitrary optical and Faraday depths. We have written a program to simulate the magnetic field structure described by shock models, and apply their solutions (corrected for typographical errors²³) cell-by-cell to calculate all four emerging Stokes parameters, which describe completely the intensity and polarization of the radiation. The numerical results discussed below are derived from this program.

Intrinsic circular polarization?

Synchrotron radiation has a small intrinsic component of circular polarization²⁵. In a uniform magnetic field, the maximum fractional

circular polarization at frequency ν is $|m_C| \approx 1/\gamma$, where γ is the Lorentz factor of electrons radiating at ν . More generally, we can write $m_C = -1.6\Lambda(\nu/\nu_{B\perp})^{-1/2}(B_u/B_{\perp})\tan\epsilon$, where B_u is the uniform component of the magnetic field, $B_{\perp}$ is the r.m.s. value of the magnetic field in the plane of the sky, $\nu_{B\perp} = eB_{\perp}/2\pi mc$ is the electron gyro frequency, and ϵ is the angle between the jet normal and the line of sight in the frame of the emitting fluid ($\cos\epsilon = \delta\sin\theta$). The factor Λ accounts for the reduction in circular polarization if there are reversals in direction of B_u , and if the charges of the radiating particles are not all the same sign. Defining $f_B = |B_u|/|B_{\perp}|$ and $f_C = (n^- - n^+)/ (n^- + n^+)$, where n^- and n^+ are the electron and positron densities respectively, then $\Lambda = f_B f_C$. The Faraday depth is also reduced by the same factor. In a real jet, even if the component of magnetic field parallel to the jet is unidirectional at its base ($f_B \approx 1$), it will decay faster with radius r ($\propto r^{-2}$) than the transverse field components ($\propto r^{-1}$). Further down the jet, the parallel component of field is likely to be dominated by sheared loops of field¹ and f_B (and hence Λ) is almost certainly small.

As the Lorentz factor of the electrons radiating at 15 GHz is $\gamma = (1.2 \times 10^3)/\eta\delta$, it is marginally possible to produce the observed circular polarization by this mechanism, but only if the uniform component of magnetic field is unidirectional, $f_B \approx 1$, which we consider physically implausible. Equally important, the numerical simulations show that the fractional circular polarization at 22 GHz should be higher than at 15 GHz by a factor of ~ 1.14 . This is because of the significant opacity at 15 GHz combined with the rather flat ($m_C \propto \nu^{-1/2}$) spectrum of intrinsic circular polarization. This predicts a circularly polarized flux at 22 GHz of 116 mJy, whereas the data give an upper limit of about 50 mJy. The argument is only slightly weaker for the inhomogeneous jet model¹⁹, for which it may be shown that $m_C \propto \nu^0$ (including opacity). For these reasons we reject intrinsic circular polarization as the origin of the observed signal. Faraday conversion of linear polarization to circular polarization, which we consider below, has the advantage of producing an intrinsically much steeper circular-polarization spectrum ($m_C \propto \nu^{-3}$ at high frequencies⁷).

Circular polarization from Faraday conversion

A less well known but equally important mechanism for generating circular polarization is Faraday conversion^{9,10}. This is based on the fact that the normal modes for radiative transfer in an anisotropic plasma are not purely circular (which leads to Faraday rotation), but slightly elliptical. The small component of linear birefringence converts Stokes parameter U to V , and vice versa. Both rotation and conversion are caused by the lowest-energy relativistic electrons, and therefore serve as a probe of the low-energy end of the electron energy distribution, which is otherwise unobservable. An important difference between them is that Faraday rotation is proportional to the electron gyro frequency, $\nu_{B\perp} = eB_{\perp}/2\pi mc$,

Table 1 July 1996 component data on 3C279

Component	r (mas)	θ (deg)	ν (GHz)	I (Jy)	m_L (%)	χ (deg)	m_C (%)
CE	...	...	15	7.58	10.7	−83	...
			22	10.46	8.7	−84	...
CW	0.10	−112	15	9.19	9.9	19	1.0
			22	10.54	12.6	27	≤0.5
K3	0.86	−122	15	0.61	8.4	−25	<2.2
			22	0.54	17.4	−8	...
K2	2.72	−119	15	0.85	2.6	55	<1.6
			22	0.53	5.8	81	...
K1	3.06	−113	15	1.67	13.7	67	<0.8
			22	1.66	13.9	67	...

Component properties determined by model fitting in the u - v plane. Component positions (r, θ) were determined from the $\nu = 22 \text{ GHz}$ data, and are given with respect to the component CE. Positions determined from the 15 GHz data are consistent with these, and show no significant shift in the position of CE with frequency. The last four columns give the total intensity (I), fractional linear polarization (m_L), position angle of the electric vectors (χ) measured from north through east, and the fractional circular polarization (m_C), respectively.

and hence to the sign of the charge on the electrons, whereas Faraday conversion is proportional to $\nu_{B\perp}^2$. Thus an equal mixture of electrons and positrons can produce Faraday conversion, but not rotation. Indeed, this is a possible observational signature of such a plasma.

It is also important to note that Faraday conversion acts on Stokes U , whereas the synchrotron mechanism produces only Stokes Q . Stokes U can be produced stochastically by a tangled field, or, more efficiently, by Faraday rotation, which we observe to be present in component CW. At small optical and Faraday depths, the fractional circular polarization produced by conversion is $m_C \approx (1/6)\tau_F\tau_C m_L^2$, where τ_F and τ_C are the Faraday depth and ‘conversion depth’ respectively, and m_L is the fractional linear polarization. For a power-law distribution of electron Lorentz factors, $n(\gamma) \propto \gamma^{-2}$ (corresponding to $\alpha = 0.5$), with a low-energy cut-off at $\gamma = \gamma_{\min}$, τ_F and τ_C can be written as⁵:

$$\tau_F \approx 1.27\tau\Lambda \left(\frac{\gamma}{\gamma_{\min}}\right)^2 \frac{\ln\gamma_{\min}}{\gamma_{\min}} \frac{B_u}{B_{\perp}} \sin \epsilon$$

$$\tau_C = -0.96\tau \ln(\gamma/\gamma_{\min})$$

where γ is the Lorentz factor of the radiating electrons.

We have explored numerically a large number of shock models to find combinations of τ_C , $\gamma_{\min}$ and Λ that reproduce the observed linear and circular polarization ($m_L = 9.9(\pm 1.0)\%$, $\tau_F = 1.1 \pm 0.4$, $m_C = 1.0(\pm 0.2)\%$). Table 2 lists the parameters of two such models, representing shocks of different strengths. In both cases, $\Lambda = 1.0$ is formally permitted, giving the largest allowed value of $\gamma_{\min}$. We have been unable to find an acceptable model in which $\gamma_{\min}$ is greater than 20. The condition $\Lambda = 1$ would require that the uniform magnetic field component, B_u , is also unidirectional, which we consider implausible for the reasons stated above. It is far more likely that $\Lambda < 1$ and $\gamma_{\min} < 20$. In fact, it is entirely possible that $\Lambda \ll 1$ and that $\gamma_{\min}$ approaches unity, though the precise value of $\gamma_{\min}$ should not be regarded as accurate, as the abrupt cut-off is artificial.

The energy flux of the jet

As outlined earlier, a value of $\gamma_{\min} < 20$ is strong evidence that the jet plasma is dominated by electron–positron pairs. The total energy flux carried by a relativistic jet can be written as²⁶:

$$F_E = 4\pi r_{\text{jet}}^2 p_{\text{jet}} \gamma_{\text{jet}}^2 \beta_{\text{jet}} c \left(1 + \frac{\gamma_{\text{jet}} - 1}{\gamma_{\text{jet}}} R\right)$$

where r_{jet} is the jet radius and p_{jet} is the total pressure. The factor $R = \rho c^2/4p_{\text{jet}}$ is the ratio of cold-matter energy density to pressure, and can be written as $R = 3(1 + 1.836f_c)/16\gamma_{\min} \ln(\gamma_{\max}/\gamma_{\min})$. A lower bound on F_E is found by assuming equipartition between the particles and the field, leading to $F_E \geq 2.5 \times 10^{45}(1 + R) \text{ erg s}^{-1}$.

Even if f_c and hence R are small, this is sufficient to power the extended (unbeamed) radio emission¹⁷, and also to produce the (presumed beamed) core radio, optical and γ -ray emission¹³, if the Doppler factor is greater than $\delta \approx 10$. Thus an electron–positron jet in 3C279 roughly matches the total observed dissipation, whereas an electron–proton jet exceeds it by one to two orders of magnitude.

Implications

Reynolds *et al.*⁶ have presented strong arguments that the well known jet in M87 is probably an electron–positron jet; they also suggested that perhaps Fanaroff–Riley²⁷ type I (FR I; low luminosity) jets are electron–positron jets and that FR II (high luminosity) jets may be electron–proton jets. We note that 3C279 (and also 3C273 and PKS0528+134) are high luminosity (type II) jets, whereas 3C84 is a low luminosity (type I) jet, suggesting that in general, extragalactic jets may be primarily electron–positron jets.

Both detailed numerical simulations of relativistic jets²⁸ and simple arguments based on ram-pressure confinement²⁹ show that jet densities must be extremely low (much less than the external density) in order to create the large diffuse lobes seen in FR II sources. This also suggests light, pair-dominated jets³⁰.

If this is the case, it points towards photon cascade^{31,32} or pion decay^{33,34} processes as the origin of most of the radiating particles, which in turn gives valuable clues to the physical mechanisms close to the central black hole that create, collimate and accelerate the jets^{1,35}. □

Methods

The VLBA is equipped with circularly polarized feeds, right (R) and left (L), at each antenna. To measure circular polarization (V) with the VLBA, we must take the difference between two large quantities, $RR = I + V$ and $LL = I - V$, and such a process is critically sensitive to gain errors.

Fringe fitting, calibration, imaging and modelling in total intensity and linear polarization were carried out using standard techniques^{16,36,37} with the software packages AIPS and DIFMAP. It is important that the polarization leakage (D) terms³⁶ at each antenna be accurately determined and removed from the RR and LL data because they induce non-closing errors that are not removed by self-calibration and could mimic a circularly polarized signal.

We have devised three different methods of making images to detect circular polarization; these methods mitigate to some extent the sensitivity to gain errors. ‘Zero V self-calibration’. The first and simplest approach is to assume there is no circularly polarized signal from the source of interest and self-calibrate with the assumption $RR = \tilde{I}_{\text{model}}$ and $LL = \tilde{I}_{\text{model}}$, where $\tilde{I}_{\text{model}}$ denotes the Fourier transform of the model total intensity distribution. This essentially forces the strongest peak in the image to appear unpolarized, and if there is no circularly polarized signal in the data, none will be induced by this procedure. For an unequal point double, resembling 3C279, circular polarization located on the strong core component will be transferred to the weak component at the same fractional level and with the reversed sign, while circular polarization originating in the weak component itself will remain largely unchanged.

Table 2 Two models for component CW of 3C279

k	ϵ (deg)	ξ	δ	$\gamma_{\min}$	Λ	m_L (%)	m_C (%)	τ_F	τ_C
0.6	20	0.3	20	15	1.00	9.2	0.89	1.38	1.37
				10	0.39	9.4	0.87	1.46	1.65
				5	0.06	9.8	0.96	1.20	2.08
				3	0.015	10.1	0.97	0.87	2.48
0.2	40	0.6	10	20	1.00	10.8	0.82	1.31	1.18
				12	0.27	10.2	0.92	1.30	1.52
				5	0.027	10.0	1.10	1.26	2.10
				3	0.006	10.5	1.00	0.83	2.44

These two models reproduce the observed polarization properties (fractional linear polarization, m_L , fractional circular polarization, m_C , and Faraday depth, τ_F) of component CW. In these models the underlying jet contains an isotropic tangled magnetic field of strength B , and a uniform component of strength $B_u = \xi B$, aligned with the jet. The shock strength is described by the compression factor, k , where unit length is compressed to length k . The line of sight makes an angle ϵ with the plane of compression, in the frame of the Doppler factor, δ is the Doppler factor, $\gamma_{\min}$ is the low-energy cut-off in the particle energy distribution, Λ accounts for magnetic field reversals and the presence of both electrons and positrons, and τ_C is the conversion depth (see text). For each model, many combinations of $\gamma_{\min}$ and Λ can reproduce the observations. As $\gamma_{\min}$ decreases, Λ must become small to avoid producing excessive Faraday rotation (τ_F).

“Phase-only mapping”. A signature of a true circularly polarized signal in a resolved source is a difference between the *RR* and *LL* closure phases. In phase-only mapping³⁸, we set the *RR* and *LL* fringe amplitudes equal to unity, but preserve their phases. For an unequal double source, this produces an antisymmetric fractional circular polarization image with half of the circularly polarized signal appearing at the position of the weaker component, and half (with opposite sign) at the corresponding position on the opposite side of the phase centre. The advantage of this is that it is almost independent of the amplitude calibration, and demonstrates that the circularly polarized signal is indeed present in the closure phases, which are preserved throughout the self-calibration procedure.

“Transfer of gains”. This depends on the excellent gain stability of the VLBA, and can resolve whether the circularly polarized signal is located on the core or in the jet. (In Homan *et al.*³⁹ we incorrectly associated the circularly polarized signal with component K1.) We assume that the amplitude gain ratio of the *R* and *L* channels at each antenna is stable over several hours and can be determined from the data on other sources in the observing run (which are interweaved with the observations of 3C279) and applied to the target source. The image shown in Fig. 1 and the results in Table 1 are derived from this approach. For 3C279, the three methods above give consistent results, and we are confident of the reality of the circularly polarized signal, and that it is in fact located in the core region.

We have run many tests and simulations on our observations and calibration techniques, to confirm the reality of our results. The circular polarization signal is present in all frequency channels and cannot be attributed to the behaviour of a single antenna. This was demonstrated by repeating the entire calibration procedure (post-fringe fit) for each frequency channel separately, and omitting each antenna in turn from the array. We checked the effect of incorrect subtraction of the antenna *D* terms from the *RR* and *LL* data by simply omitting the correction altogether. This increased the noise on the *V* image but did not generate a significant spurious circularly polarized component.

Finally, we created three model sources, similar to 3C279 in *I* and *P* structure, with no initial *V*, *V* on the core, and *V* on the jet component. These models contained thermal noise, *D*-term leakage, time-dependent complex gains, and baseline-based errors. We carried these models through our standard (post fringe-fit) calibration procedure and found that we were able to detect a circularly polarized signal placed in the model data. Equally important, we were unable to generate a *V* signal when the model contained no *V*.

Received 2 April; accepted 28 July 1998.

1. Begelman, M. C., Blandford, R. D. & Rees, M. J. Theory of extragalactic radio sources. *Rev. Mod. Phys.* **56**, 255–351 (1984).
2. Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
3. Königl, A. Relativistic jets as x-ray and gamma ray sources. *Astrophys. J.* **243**, 700–709 (1981).
4. Sikora, M., Begelman, M. C. & Rees, M. J. Comptonization of diffuse ambient radiation by a relativistic jet. The source of gamma rays from blazars? *Astrophys. J.* **421**, 153–162 (1994).
5. Celotti, A. & Fabian, A. C. The kinematic power and luminosity of parsecscale radio jets—an argument for heavy jets. *Mon. Not. R. Astron. Soc.* **264**, 228–236 (1993).
6. Reynolds, C. S., Fabian, A. C., Celotti, A. & Rees, M. J. The matter content of the jet in M87: evidence for an electron-positron jet. *Mon. Not. R. Astron. Soc.* **283**, 873–880 (1996).
7. Wardle, J. F. C. Upper limits on the Faraday rotation in variable radio sources. *Nature* **269**, 563–566 (1977).

8. Jones, T. W. & O’Dell, S. L. Physical conditions in polarized compact radio sources. *Astron. Astrophys.* **61**, 291–293 (1977).
9. Jones, T. W. & O’Dell, S. L. Transfer of polarized radiation in self-absorbed synchrotron sources. I—Results for a homogeneous source. *Astrophys. J.* **214**, 522–539 (1977).
10. Jones, T. W. Polarization as a probe of magnetic field and plasma properties of compact radio sources. *Astrophys. J.* **332**, 678–695 (1988).
11. Burbidge, E. M. & Rosenberg, F. D. The redshift of the quasi-stellar radio source 3C279. *Astrophys. J.* **142**, 1673–1674 (1965).
12. Wehrle, A. *et al.* Multiwavelength observations of a dramatic high-energy flare in the blazar 3C279. *Astrophys. J.* **497**, 178–187 (1998).
13. Hartman, R. C. *et al.* Simultaneous multiwavelength spectrum and variability of 3C279 from 10⁹ Hz to 10²⁴ Hz. *Astrophys. J.* **461**, 698–712 (1996).
14. Unwin, S. C., Cohen, M. H., Biretta, J. A., Hodges, M. W. & Zensus, J. A. Superluminal motion in the quasar 3C279. *Astrophys. J.* **340**, 117–128 (1989).
15. Carrara, E. A., Abraham, Z., Unwin, S. C. & Zensus, J. A. The millisecond structure of the quasar 3C279. *Astron. Astrophys.* **279**, 83–89 (1993).
16. Leppänen, K. J., Zensus, J. A. & Diamond, P. J. Linear polarization imaging with Very Long Baseline Interferometry at high frequencies. *Astron. J.* **110**, 2479–2492 (1995).
17. de Pater, I. & Perley, R. A. The radio structure of 3C279. *Astrophys. J.* **273**, 64–69 (1983).
18. Cotton, W. D. *et al.* 3C279—the case for superluminal expansion. *Astrophys. J.* **229**, L115–L117 (1979).
19. Blandford, R. D. & Königl, A. Relativistic jets as compact radio sources. *Astrophys. J.* **232**, 34–48 (1979).
20. Marscher, A. P. Accurate formulae for the self-Compton x-ray flux density from a uniform spherical compact radio source. *Astrophys. J.* **264**, 296–297 (1983).
21. Ghisellini, G., Padovani, P., Celotti, A. & Maraschi, L. Relativistic bulk motion in active galactic nuclei. *Astrophys. J.* **407**, 65–82 (1993).
22. Marscher, A. P. & Gear, W. K. Models for high-frequency radio outbursts in extragalactic sources, with application to the early 1983 millimeter-to-infrared flare of 3C273. *Astrophys. J.* **298**, 114–127 (1985).
23. Hughes, P. A., Aller, H. D. & Aller, M. F. Synchrotron emission from shocked relativistic jets. I—The theory of radio-wavelength variability and its relation to superluminal motion. *Astrophys. J.* **341**, 79–99 (1989).
24. Wardle, J. F. C., Cawthorne, T. V., Roberts, D. H. & Brown, L. F. Interpretation of VLBI kinematic and polarization data: application to 3C345. *Astrophys. J.* **437**, 122–135 (1994).
25. Lee, M. I. C. & Westfold, K. C. Elliptic polarization of synchrotron radiation. *Astrophys. J.* **154**, 499–514 (1968).
26. Bicknell, G. V. in *Energy Transport in Radio Galaxies and Quasars* (eds Hardee, P. E., Bridle, A. H. & Zensus, J. A.) 253–260 (ASP Conf. Ser. 100, Astronomical Soc. of the Pacific, San Francisco, 1996).
27. Fanaroff, B. & Riley, J. M. The morphology of extragalactic radio sources of high and low luminosities. *Mon. Not. R. Astron. Soc.* **167**, 31P–36P (1974).
28. Duncan, G. C. & Hughes, P. A. Simulations of relativistic extragalactic jets. *Astrophys. J.* **436**, L119–L122 (1994).
29. Alexander, P. & Pooley, G. G. in *Cygnus A—Study of a Radio Galaxy* (eds Carilli, C. L. & Harris, D. E.) 149–157 (Cambridge Univ. Press, 1996).
30. Scheuer, P. A. G. in *Energy Transport in Radio Galaxies and Quasars* (eds Hardee, P. E., Bridle, A. H. & Zensus, J. A.) 1–7 (ASP Conf. Ser. 100, Astronomical Soc. of the Pacific, San Francisco, 1996).
31. Blandford, R. D. & Levinson, A. Pair cascades in extragalactic jets. I: Gamma rays. *Astrophys. J.* **441**, 79–95 (1995).
32. Marcowith, A., Henri, G. & Pelletier, G. Gamma-ray emission of blazars by a relativistic electron-positron beam. *Mon. Not. R. Astron. Soc.* **277**, 681–699 (1995).
33. Mannheim, K. & Biermann, P. L. Gamma-ray flaring of 3C279—A proton-initiated cascade in the jet? *Astron. Astrophys.* **253**, L21–L24 (1992).
34. Bednarek, W. & Calvani, M. X- and gamma-ray emission from 3C273. *Astron. Astrophys.* **245**, 41–47 (1991).
35. Burns, M. L. & Lovelace, R. V. E. Theory of electron-positron showers in double radio sources. *J.* **262**, 87–99 (1982).
36. Roberts, D. H., Wardle, J. F. C. & Brown, L. F. Linear polarization radio imaging at millisecond resolution. *Astrophys. J.* **427**, 718–744 (1994).
37. Cotton, W. D. Calibration and imaging of polarization sensitive Very Long Baseline Interferometer observations. *Astron. J.* **106**, 1241–1248 (1993).
38. Wardle, J. F. C. & Roberts, D. H. in *Compact Extragalactic Radio Sources* (eds Zensus, J. A. & Kellermann, K. I.) 217–222 (Workshop Proc., NRAO, Socorro, NM, 1994).
39. Homan, D. C., Ojha, R., Wardle, J. F. C. & Roberts, D. H. in *Radio Emission from Galactic and Extragalactic Compact Sources* (eds Zensus, J. A., Taylor, G. B. & Wrobel, J. M.) (IAU Colloq. 164, Astronomical Soc. of the Pacific, San Francisco, in the press).

Acknowledgements. We thank T. Jones for reading the manuscript. This work was supported by NSF and NASA.

Correspondence and requests for materials should be addressed to J.F.C.W. (e-mail: ifcw@quasar.astro.brandeis.edu).

The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM

J. Sayos*, C. Wu*, M. Morra*, N. Wang*, X. Zhang*, D. Allen*, S. van Schaik*, L. Notarangelo†, R. Geha‡, M. G. Roncarolo§, H. Oettgen‡, J. E. De Vries||, G. Aversa|| & C. Terhorst*

* Division of Immunology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA

† Department of Pediatrics, University of Brescia, Brescia 25125, Italy

‡ Division of Immunology, The Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

§ Cellular Therapy Laboratory, Telethon Institute for Gene Therapy, Milan 20132, Italy

|| Novartis Forschungsinstitut Gesellschaft mbH, Vienna A-1235, Austria

In addition to triggering the activation of B- or T-cell antigen receptors, the binding of a ligand to its receptor at the cell surface can sometimes determine the physiological outcome of interactions between antigen-presenting cells, T and B lymphocytes. The protein SLAM (also known as CDw150), which is present on the surface of B and T cells, forms such a receptor–ligand pair as it is a self-ligand. We now show that a T-cell-specific, SLAM-associated protein (SAP), which contains an SH2 domain and a short tail, acts as an inhibitor by blocking recruitment of the SH2-domain-containing signal-transduction molecule SHP-2 to a docking site in the SLAM cytoplasmic region. The gene encoding SAP maps to the same area of the X chromosome as the locus for X-linked lymphoproliferative disease (XLP) and we found mutations in the SAP gene in three XLP patients. Absence of the inhibitor SAP in XLP patients affects T/B-cell interactions induced by SLAM, leading to an inability to control B-cell proliferation caused by Epstein–Barr virus infections.

SLAM (for signalling lymphocyte-activation molecule), is a glycosylated type-I transmembrane protein, of relative molecular mass 70K, that is present on the surface of B and T cells; it is a high-affinity self-ligand. As triggering of SLAM co-activates T- or B-lymphocyte responses, it is considered to be important in bidirectional T ↔ B-cell stimulation^{1–7}. Monoclonal antibodies directed against SLAM induce proliferation of human CD45RO^{high} memory T cells in a CD28-independent manner, whereas resting T cells are unresponsive^{1–3}. Anti-SLAM acts as a co-activator, together with antibodies against the T-cell antigen receptor (TCR) or complexes of peptides with the major histocompatibility complex (MHC) on the surface of antigen-presenting cells, in stimulating resting T cells to synthesize DNA and/or to initiate cytokine gene activation cascades^{1,2}. Triggering of signals by anti-SLAM antibodies induces the interferon- γ gene and redirects Th2 responses of antigen-specific T-cell clones to a Th1 or Th0 phenotype^{3–5}.

Exposure of B cells to a soluble form of the SLAM ectodomain or to L cells transfected with membrane-bound SLAM dramatically enhances B-cell proliferation and production of IgM, IgG and IgA, in conjunction with anti-CD40 monoclonal antibodies or other co-stimuli⁶. Anti-SLAM antibodies alone fail to enhance B-cell proliferation, even having an inhibitory effect⁶. Thus, SLAM-induced signal-transduction events in T lymphocytes are different from those in B cells.

Cloning of SAP

In order to determine the mechanism by which SLAM induces signalling, we identified proteins that interact with the cytoplasmic domain of human SLAM. Using a yeast two-hybrid system with the 77-amino-acid human SLAM cytoplasmic domain as bait and a human T-cell library, eight complementary DNA clones were isolated that encoded a 128-amino-acid polypeptide chain which we term SLAM-associated protein, or SAP. To confirm that the cDNAs were full-length, we also isolated cDNA clones from a

human T-cell (Jurkat) library and from a murine (C57B16) thymus library. The predicted human SAP protein comprises a single SH2 domain (residues 6–102) followed by a short 26-amino-acid tail (residues 103–128). Human and murine SAP are highly homologous (96%) in both the SH2 and tail domains (Fig. 1). The SAP SH2 domain was clearly related to other SH2 domains, particularly those of SHIP (45%), EAT-2 (40%) and Abl (35%), but its short tail was unlike any known protein domain.

Antibodies directed against human or mouse SAP-tail sequences detected a 15K protein in detergent lysates made from either human T cells (Fig. 2a) or murine thymocytes (Fig. 2b). Because the observed relative molecular mass was consistent with the predicted value, our human and mouse SAP cDNAs are unlikely to code for a fragment of a larger protein. We found that SLAM and SAP interact in T lymphocytes because a human SAP fusion protein, with glutathione-S-transferase, can specifically precipitate SLAM from detergent lysates made from a murine T-cell line (EL-4) transfected with human SLAM. By contrast, no interaction was detected in lysates made with untransfected EL-4 cells (Fig. 2c). Conversely, SAP was co-immunoprecipitated with anti-SLAM antibodies using detergent lysates made from activated human peripheral blood lymphocytes⁵ (Fig. 2d).

Identification of SAP as a T-cell protein

Human SAP messenger RNA expression was highest in the thymus and less in spleen and peripheral blood lymphocytes (Fig. 3a). SAP was expressed in all major subsets of human T cells (CD4⁺, CD45RO⁺, CD45RA⁺ and CD8⁺) (Fig. 3c), in the T-cell tumour cell line Jurkat, and in the Burkitt lymphoma line Raji which is positive in the Epstein–Barr virus (EBV). No transcripts were detected in the EBV⁺ Burkitt lymphomas Namalwa or BL41/B95-8, the EBV-transformed cell line X50-7, the EBV⁺ marmoset cell lines B95-8 and FF41, the EBV⁺ Burkitt lymphoma BJAB, or in the pre-B ALL lines LAZ 221 and NALM 6 (Fig. 3, and data not shown).

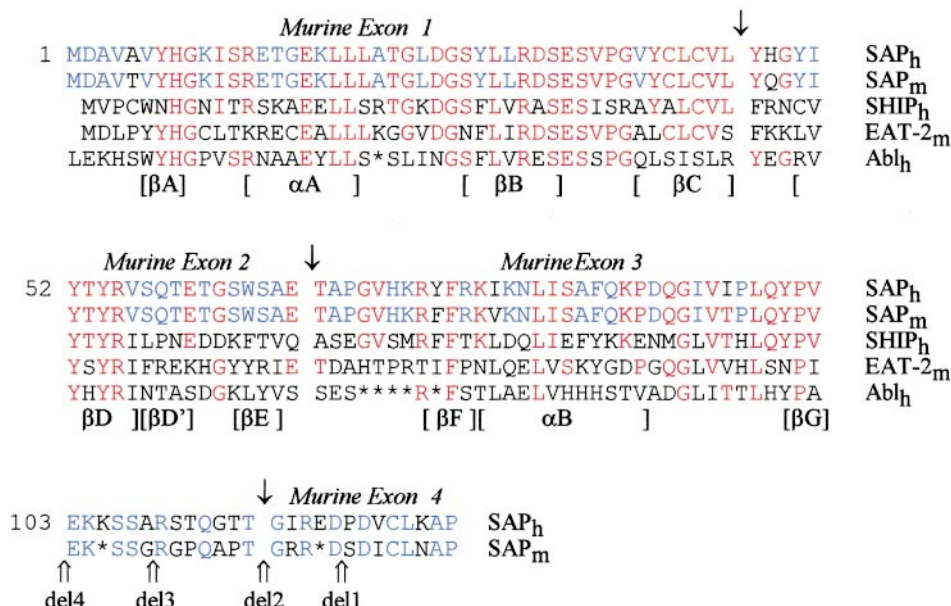


Figure 1 Comparison of deduced amino-acid sequences of human SAP, murine SAP and other SH2-domain-containing proteins. Amino acids that are only found in human and murine SAP are in blue; amino acids found in other SH2 domains are in red. Consensus α -helices and β -sheets are indicated in brackets. Exon/intron boundaries as derived from the murine SAP gene (C.W., J.S. and C.T., manuscript in preparation) are demarcated by downward arrows. Four truncation mutants were generated in the tail of human SAP (del 1–4) and their locations in

the amino-acid sequence are shown (upward arrows). Comparisons are made between SH2 domains that were found to be most similar in a computer-aided search, namely human SHIP, murine EAT-2 and human Abl. A computer-aided search for the SAP cDNA in the EST library identified three human clones (accession numbers N89899, W19453 (fetal lung) and AA354319 (Jurkat)) and one highly homologous murine T-cell-derived clone of 294 bp (accession number AA255258).

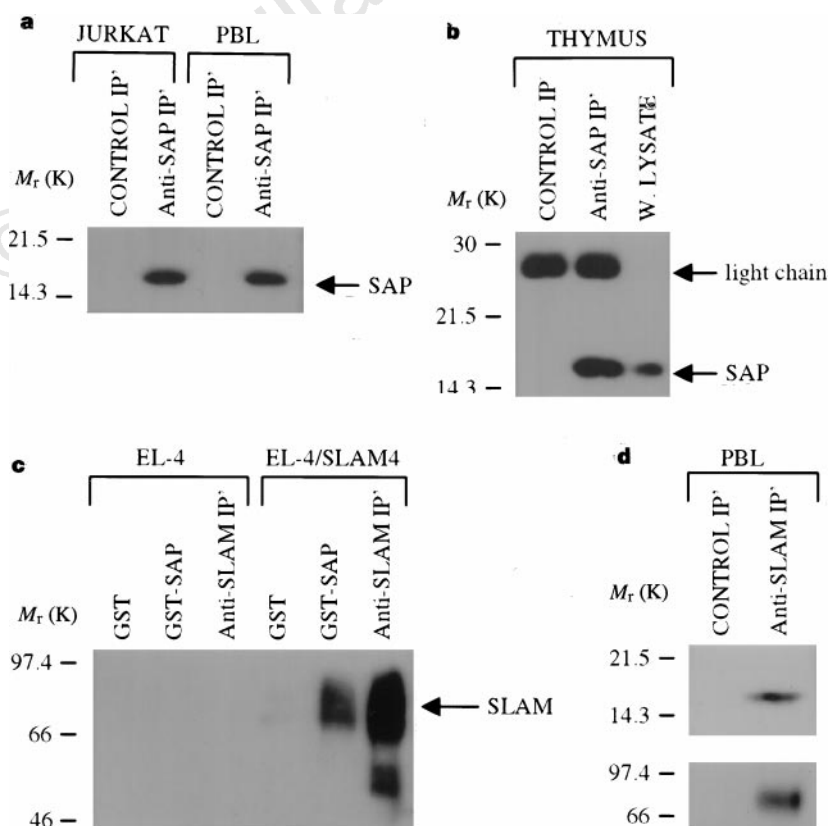


Figure 2 Characterization of the human and murine SAP proteins and their interactions with SLAM. **a**, Jurkat cells and human peripheral blood lymphocytes (PBL) were lysed and postnuclear lysates (1 mg ml⁻¹) were immunoprecipitated (IP) with a rabbit antiserum directed against human SAP or with a control serum (3 μ l). Samples were analysed on SDS-PAGE and immunoblotted with anti-human-SAP (1/1,000). **b**, Postnuclear lysates (1 mg ml⁻¹) from C57B1/6 thymocytes were immunoprecipitated with an anti-murine SAP antibody (3 μ l), resolved on SDS-PAGE and immunoblotted with anti-mouse-SAP (1/1,000). **c**, SLAM- or vector-transfected EL-4 cells were cell-surface biotinylated and lysed in

detergent. Postnuclear lysates (1 mg ml⁻¹) were incubated with 5 μ g GST or GST-SAP for 1 h in the presence of glutathione beads or 1 μ g anti-SLAM antibodies (a gift from DNAX Research Institute) for 3 h in the presence of protein G beads. Bead-associated proteins were extracted and immunoblotted with streptavidin-HRP. **d**, Human PBL were activated with phytohaemagglutinin for 5 d. Cells were lysed (1 $\times$ 10⁶ per ml) and immunoprecipitated with an anti-SLAM monoclonal antibody or with a control antibody. Immunoprecipitates were resolved on SDS-PAGE and immunoblotted with anti-SLAM serum (top) or with a rabbit anti-SAP antibody (bottom).

A small amount of SAP mRNA detected in the small intestine probably resulted from lymphocyte contamination. In wild-type mice, SAP is expressed in T cells but not in B cells, whereas no mRNA or protein could be detected in CD3 ϵ^{null} mice which lack T cells (C.W., J.S. and C.T., manuscript in preparation). Both human mRNA species encoding SAP (2.5 and 0.9 kilobases (kb)) (Fig. 3) were represented among the cDNA clones isolated in the two-hybrid system and encoded the same open reading frame, but differed in their 3' untranslated sequences. Collectively, our results show that, unlike SLAM, SAP is primarily expressed in T cells.

SAP is encoded by the XLP disease gene

By using a 45-kb pBAC clone that contained all four exons of murine SAP (Fig. 1), the SAP gene was localized within band A5.1 of the murine X chromosome (C.W., J.S. and C.T., manuscript in preparation). Because of synteny between murine band A5.1 and

human Xq25, where the locus for the immune deficiency X-linked lymphoproliferative disease (XLP) had been mapped⁸⁻¹¹, we investigated whether the SAP gene could be involved in this disease. Moreover, the observation that in XLP uncontrolled B-cell proliferation following EBV infection often leads to fatal infectious mononucleosis, malignant lymphomas, hypogammaglobulinaemia or aplastic anaemia^{11,12} was in agreement with the involvement of SLAM in T $\leftrightarrow$ B-cell interactions. To test the idea that SAP is encoded by the XLP gene, we isolated SAP cDNAs from peripheral blood mononuclear cells of three XLP patients: a previously described patient, A1 (ref. 13), and two brothers, B1 and B2, who were recently diagnosed as having XLP.

XLP patient A1 developed hypogammaglobulinaemia and recurrent pulmonary infections a few months after severe EBV-induced infectious mononucleosis¹³. His family history and the persistence of an unbalanced virus-host relationship after EBV infection are consistent with the main features of XLP syndrome¹³. When the products from reverse transcription with polymerase chain reaction (RT-PCR) of T cells from patient A1 were analysed by polyacrylamide gel electrophoresis (Fig. 4a), three products were initially found of 629 base pairs (bp) (the full-length SAP coding sequence), 565 bp and 520 bp. On cloning the RT-PCR products of patient A1, four different mRNAs were identified: namely, full-length hSAP; a coding sequence, Δ E2, which lacked the 64 nucleotides of exon 2 (A1-1 in Fig. 5); E3 Δ 55, with a 55-nucleotide deletion in exon 3 (hSAP Δ 55 in Fig. 5); and a cDNA with both a deleted exon 2 and the deletion E3 Δ 55 (A1-2 in Fig. 5). T cells from patient A1 thus had four mRNA species coding for different forms of SAP, of which two were detected in healthy individuals (629 and 574 bp) (Fig. 5) and two were specific for the patient (A1-1 of 565 bp and A1-2 of 520 bp; note that the 574- and 565-bp species cannot be separated by PAGE in Fig. 4). The predicted amino-acid sequences were indicative of two truncated proteins, in which essentially only the first exon of SAP is intact.

RT-PCR products from peripheral blood mononuclear cells of 60

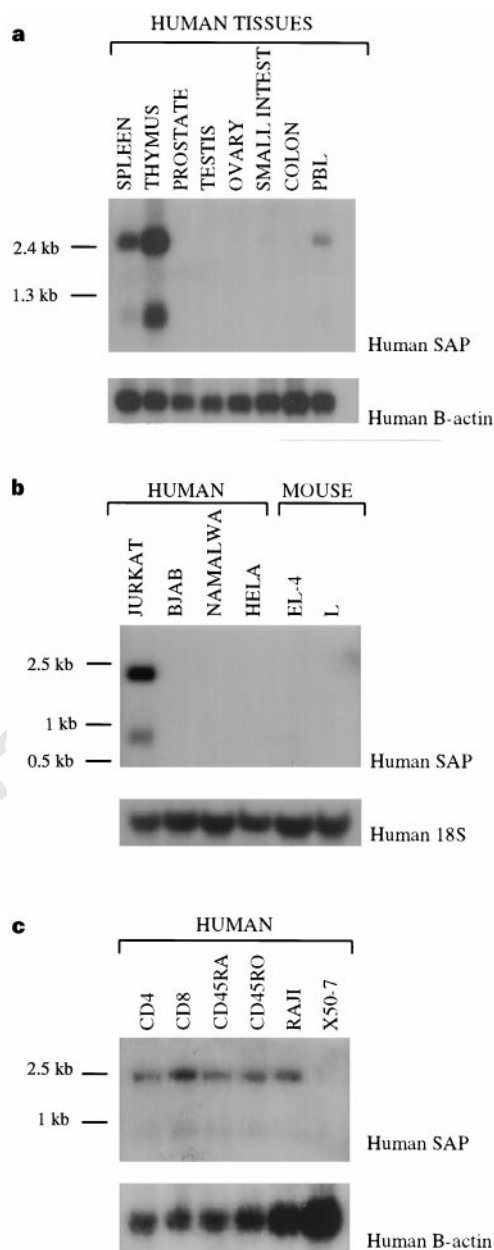


Figure 3 SAP expression. Northern blot analysis of **a**, poly(A)⁺ mRNAs isolated from various human tissues (Clontech); **b**, total RNAs from different human and mouse cell lines (20 μ g per lane); or **c**, human T-cell subsets and two EBV⁺ B-cell lines. Specific RNAs were detected using a ³²P-radiolabelled human SAP probe; β -actin or 18S rRNA probes were used as loading controls.

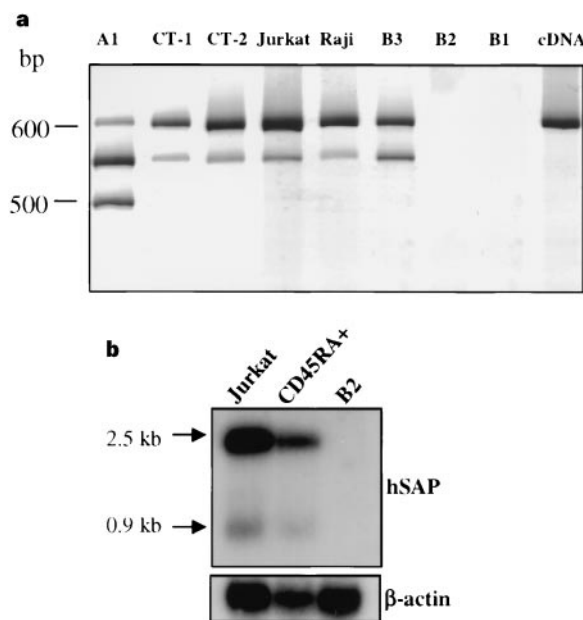


Figure 4 mRNA analysis of XLP patients. **a**, Complementary DNA was isolated from peripheral blood mononucleocytes (PBMC) from patients and healthy donors by RT-PCR. DNA products were resolved on 10% polyacrylamide gels. Samples A1, B1 and B2 are from XLP patients; sample B3 is from a healthy brother of B1 and B2; samples CT-1 and CT-2 are from healthy controls. **b**, Northern blot analysis of total RNAs from Jurkat T cells, from a subset of CD45RA^{high} cells from a healthy donor, and from PBMC from patient B2 (20 μ g per lane). Specific RNAs were detected using a ³²P-radiolabelled human SAP probe; a β -actin probe was used as loading control.

healthy individuals contained two mRNA species (of 629 and 574 bp), as judged by PAGE and nucleotide-sequence analysis (Figs 4a, 5). These represented hSAP and E3Δ55 (hSAPΔ55 in Fig. 5a) in which part of exon 3 had been deleted. The E3Δ55 coding sequence started at the beginning of exon 3 at nucleotide position 288 and ended at nucleotide 342 (Fig. 5a). Because of a frameshift, the predicted protein sequence was identical to that of the first two exons of SAP (amino acids 1–67), followed by a nine-amino-acid sequence and a stop codon (Fig. 5b). The results taken together indicate that the mRNAs of patient A1 had a deleted exon 2 or a deletion in exon 2 plus part of exon 3. These mutant forms derived from patient A1 were estimated to represent 90% of his SAP mRNA.

The 55-nucleotide deletion in exon 3 in healthy individuals is readily explained by the presence of an infrequently used splicing acceptor site within that exon: CAG preceded by a CT-rich stretch of nucleotides¹⁴ (Fig. 6a). But to explain the frequent deletion of exon 2 in patient A1, the intron sequences surrounding exons 1, 2 and 3 had to be determined, so we designed a PCR assay based on sequences of at least 150 nucleotides of all four introns. This was facilitated by a new sequence in the EMBL database (accession number AL022718) indicating that *XLP* encoded a protein, SH2DIA. Its cDNA, genomic DNA and protein sequences were identical to human SAP, and in addition the exon/intron bound-

aries reported were identical to those in the mouse gene (Fig. 1). Upon sequencing the 443-bp PCR product derived from the exon-2 area, we found a C → G mutation in the nucleotide adjacent to the exon 2 splice acceptor site in the DNA of patient A1 (Fig. 6a). This mutation is known to affect the efficiency of the splice acceptor site¹⁵ and explains the skipping of exon 2 in most (~90%) of the patient's SAP mRNA species.

To exclude the possibility that the point mutation found in XLP patient A1 represented a genetic polymorphism, we designed a restriction enzyme assay based on the fact that the C → G mutation generated a novel *MnII* restriction site in the 443-bp exon-2 PCR product. As shown in Fig. 6b, *MnII* generated four fragments (of 187, 151, 64 and 41 bp) with A1's 443-bp PCR product, whereas DNA samples from 108 healthy individuals generated three fragments (of 251, 151 and 41 bp) (see, for example, CT-1, CT-2 and B3 in Fig. 6b, and results not shown). Because patient A1 was Italian, these controls included DNA from 50 healthy Italian females (that is, 100 X chromosomes). The remaining 86 X chromosomes were from individuals with a random genetic background in the Boston area (28 female and 30 male). Collectively, our results demonstrate that A1 has a mutation affecting the exon 2 splice acceptor site, which gives rise to severely truncated forms of SAP. As this mutation was not detected in 186 X chromosomes from healthy individuals, it

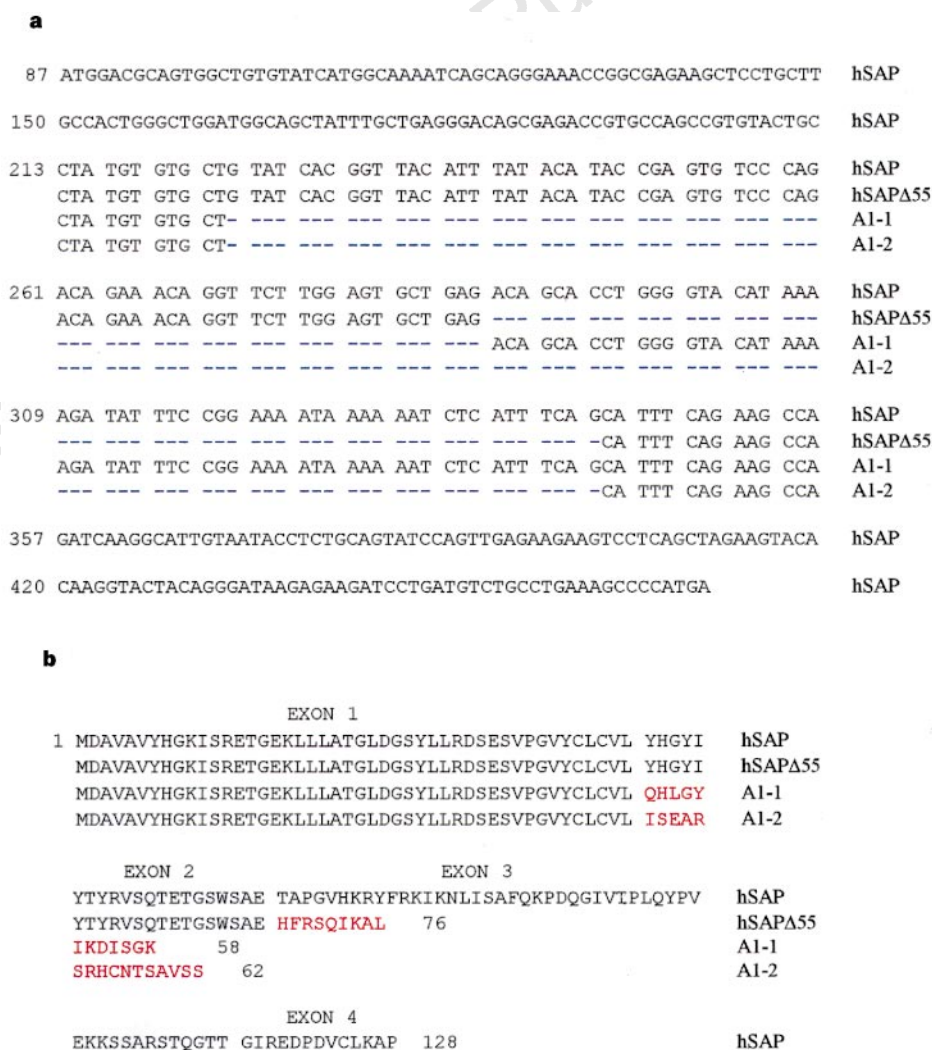


Figure 5 Nucleotide and amino-acid sequences of SAP isolated from a patient with XLP. **a**, Comparison of the nucleotide sequence of two cDNAs isolated by RT-PCR from an XLP patient with those of human SAP (hSAP and hSAPΔ55). **b**, Comparison of the predicted amino-acid sequence of SAP from an XLP patient

with that of normal human SAP. DNA products were cloned in the pCR2.1 vector and the nucleotide sequence determined from two cDNA clones in both directions in an ABI prism 377 DNA sequencer.

did not represent a genetic polymorphism. These results indicate that SAP is encoded by the XLP gene.

The two other XLP patients, brothers B1 and B2, had a deletion of the SAP gene, whereas a healthy brother, B3, had a normal SAP gene. B1 was a 23-year-old male with a history of recurrent pulmonary infections, dysgammaglobulinaemia (raised IgA and IgM), poor specific antibody responses to tetanus toxoid antigen, and a depressed T-cell proliferative response to mitogens. He developed fever, pneumonia and hilar adenopathy which quickly progressed to fulminant infectious mononucleosis with haemophagocytosis; EBV infection was documented by PCR and serology. B2 also suffers from XLP syndrome, which manifests as pancytopenia, splenomegaly, dysgammaglobulinaemia and depressed T-cell proliferative responses to mitogens. B3 appears to be healthy, but two other brothers, for whom no material was available, had classic XLP symptoms: one has recurrent EBV⁺ non-Hodgkin's lymphomas and another died with a diagnosis of Wegener's granulomatosis and pulmonary infiltration.

In contrast to the healthy sibling B3, no RT-PCR product encoding hSAP could be isolated from the peripheral blood mononuclear cells of patients B1 and B2 (Fig. 4a). This was because no mRNA was detected in northern blotting experiments with RNA from patients B1 (data not shown) and B2 (Fig. 4b). Thus, either the SAP gene was not transcribed in T cells from B1 and B2 or the gene was deleted, which commonly happens in XLP patients^{11,12}. As none of the four exons encoding SAP could be generated in a PCR analysis based upon genomic DNA from B1 and B2 (Fig. 6c) and because a control PCR reaction for exon 2 of *BRCA1* gave the expected product, we conclude that the SAP gene was deleted in XLP patients B1 and B2. By contrast, DNA from B3 generated PCR fragments of the expected size and with the wild-type sequence (a similar result has been obtained by J. Sumegi, personal communication). These analyses of XLP patients A1, B1 and B2 show that the XLP gene encodes SAP.

Binding of the SH2 domain of SAP

Identification of SAP as the gene product that is altered in XLP raised the question of how this SLAM-associated T-cell protein could account for the immunological disturbances of the disease. The SLAM cytoplasmic domain contains three tyrosine residues (Y281, Y307 and Y327) that are surrounded by consensus SH2-domain-binding sequences¹, which suggests a possible interaction with the SH2 domain of SAP. Because SAP binds to SLAM in the yeast two-hybrid system, the binding of the SAP SH2 domain to SLAM must be more complex than the usual interactions with a short phosphotyrosine-containing peptide.

To determine to which portion of the cytoplasmic domain of SLAM the SAP SH2 domain binds, we co-expressed constructs encoding the CD8 extracellular and transmembrane domains and truncated cytoplasmic domains of SLAM with human SAP in COS-7 cells (Fig. 7a). Both CD8-SLAM proteins, that is CD8-SLAM3 (derived from a natural variant of SLAM with a short cytoplasmic domain containing only Y281) and CD8-SLAMdel1 (containing Y281 and Y307), co-precipitated SAP with the same efficiency as SLAM itself (Fig. 7a); two control chimaeric proteins, comprising the CD8 transmembrane and ecto domains and either the CD3- ϵ or the CD3- ζ cytoplasmic domain, did not co-precipitate SAP (Fig. 7a). This suggested the presence of a specific SAP-binding site around the most membrane-proximal tyrosine residue (Y281) of the SLAM cytoplasmic tail. As predicted from the *in vivo* results, only a peptide (SLAM Y1) containing Y281 coupled to agarose beads was able specifically to precipitate SAP from murine thymocyte lysates, and not peptides containing the other tyrosine residues in the cytoplasmic domain of SLAM (Fig. 7b). More important, binding of SAP to agarose-coupled SLAM Y1 was blocked only by the same peptide regardless of whether Y281 was replaced by phenylalanine (SLAM F1) or by a peptide in which phosphotyrosine replaced Y281 (SLAM pY1) (Fig. 7c). This confined the binding site to a short amino-acid sequence in SLAM and confirmed that

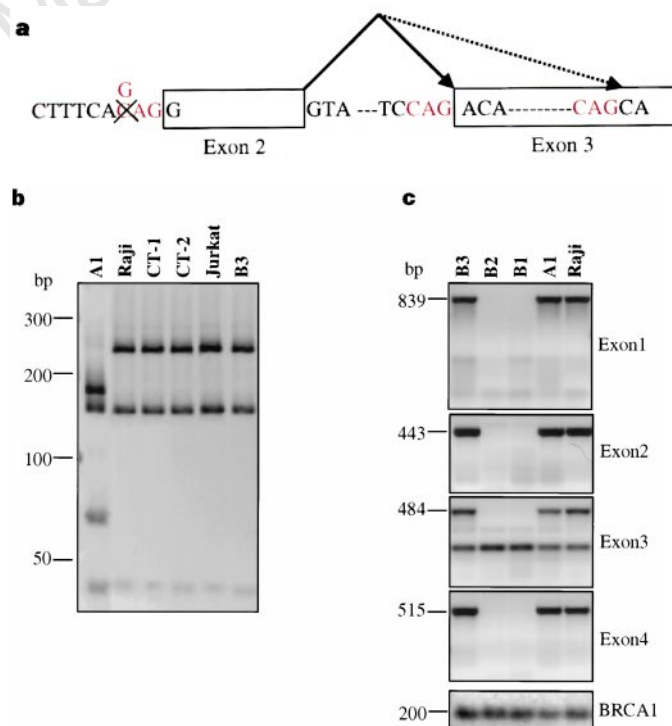


Figure 6 Genomic analysis of patients with XLP. **a**, Location of the point mutation near exon 2 of patient A1; a mechanism is shown to explain the generation of the variant from hSAP Δ 55 found in all healthy individuals. **b**, Exon 2 sequence was amplified by PCR from genomic DNA from patient A1 and from three healthy individuals (CT-1, CT-2 and B3) and from two cell lines (Raji and Jurkat). In addition, PCR products from 78 healthy women and 30 healthy men were analysed (data

not shown). DNA products were digested with *MnII* and resolved on a 10% polyacrylamide gel by standard methods. **c**, hSAP exon 1, exon 2, exon 3 and exon 4 and *BRCA1* exon 2 were amplified by PCR from genomic DNA from patients A1, B1, B2 and B3, and from the cell line Raji. DNA products were resolved on a 2% agarose gel.

phosphorylation of Y281 was not required for binding of the two polypeptide chains. However, we found that the SH2 domain of SAP was functionally intact as it could be purified on a Sepharose-phosphotyrosine column (data not shown).

In addition, four SAP mutants with deletions in the tail domain (Fig. 1) and co-transfected with the CD8-SLAM3 chimera were able to interact with the short cytoplasmic domain of SLAM3, which contained the Y281 segment (Fig. 7d). Collectively these results show that SLAM interacts with SAP through part of its SH2 domain.

SAP blocks recruitment of SHP-2 to SLAM

The novelty of SLAM-SAP recognition, in which the usual requirements for SH2-domain interactions seem to be altered, suggested that SAP could block physiological reactions involving SLAM, acting as a natural inhibitor. This model would predict that SAP mutants that are unable to bind to SLAM could not act as inhibitors. We were able to test this hypothesis because we previously found (J.S. *et al.*, manuscript in preparation) that after phosphorylation by *c-fyn*, SLAM recruited the tyrosine phosphatase SHP-2.

First, two SAP mutants were tested: the A1 deletion (Δ exon2 + E3 Δ 55) and a classic mutation in the SH2 domain that affects binding of phosphotyrosine (SAP32R $\rightarrow$ Q). Neither

mutant bound to the cytoplasmic tail of SLAM, as judged by *in vivo* (Fig. 8a) or *in vitro* (Fig. 8b) assays. Thus, neither SAP mutant could act as a dominant-negative mutant through SLAM. More important, the experiment with mutant SAP32R $\rightarrow$ Q revealed that the interaction of SAP and SLAM depends on an intact phosphotyrosine binding pocket. An XLP patient with a mutation in R32 has recently been found (J. Sumegi, personal communication). These results, and the failure of the two SAP mutants to bind in patients, further support the idea that SAP-SLAM interaction is defective in XLP.

Next, SLAM and *c-fyn* were co-transfected with or without SAP into COS-7 cells, which contain endogenous SHP-2. If SLAM was phosphorylated by *c-fyn*, SHP-2 was co-precipitated with anti-SLAM antibody (Fig. 8c); in the absence of *c-fyn*, SLAM did not bind SHP-2. Introduction of SAP, together with SLAM and *c-fyn*, blocked the binding of SHP-2 and SLAM. The absence of SHP-2 in the presence of SAP could not be due to lesser phosphorylation of SLAM, because a significantly larger percentage of SLAM molecules was phosphorylated in the presence of SAP than in its absence. These results were confirmed by reciprocal co-immunoprecipitation with anti-SAP reagents (data not shown). By contrast (Fig. 8d), mutant SAP32R $\rightarrow$ Q did not prevent binding of SHP-2 to SLAM. A similar result was obtained with the A1 deletion mutant (data not

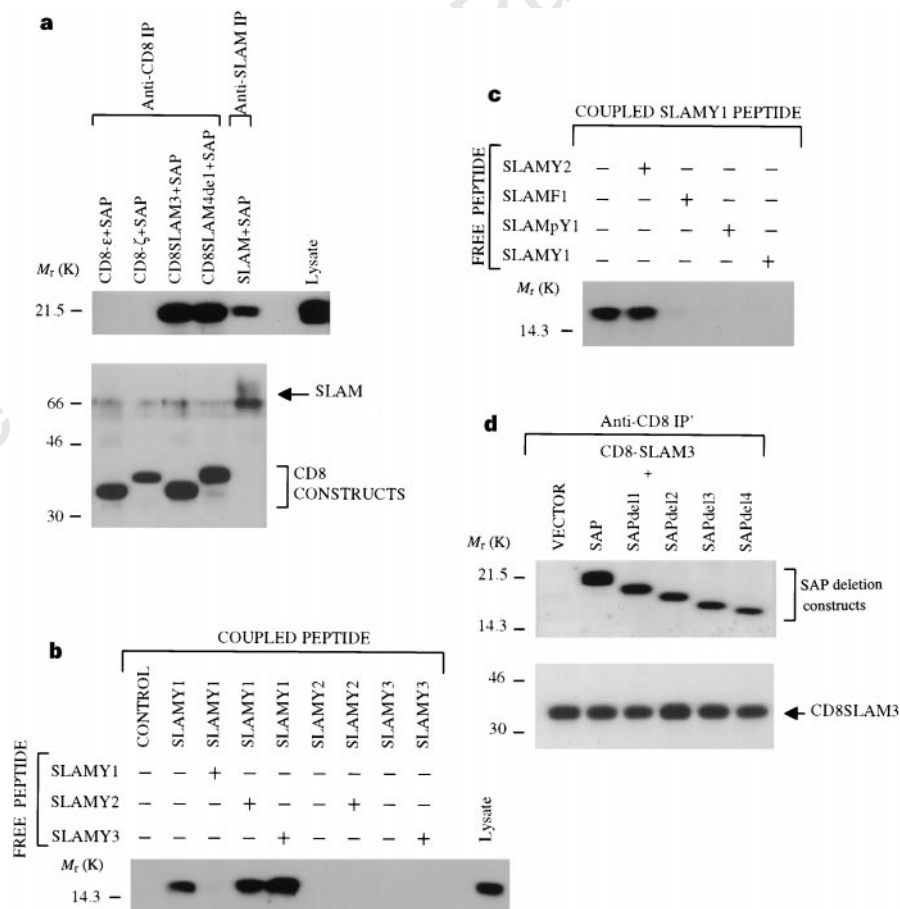


Figure 7 The SAP SH2 domain binds to a specific sequence in the cytoplasmic domain of SLAM. **a**, COS-7 cells were transfected, biotinylated and lysed. Postnuclear lysates were immunoprecipitated with antibodies anti-CD8 (OKT8) or anti-SLAM (ref. 1) and immunoprecipitates were immunoblotted with anti-human-SAP rabbit sera (top panel) or streptavidin (bottom). **b**, Postnuclear mouse thymocyte lysates (1 mg ml⁻¹) were incubated for 1 h with different peptides coupled to beads (7 μ M) or with control beads in the absence or presence of free peptides (280 μ M). Bead-associated proteins were extracted and immunoblotted with an anti-human-SAP rabbit serum (1/1,000). **c**, Postnuclear mouse

thymocyte lysates (1 mg ml⁻¹) were incubated for 1 h with SLAM Y1 peptide coupled to beads (7 μ M) in the absence or presence of free peptides (70 μ M). Bead-associated proteins were extracted and immunoblotted with an anti-murine-SAP rabbit serum (1/1,000). **d**, COS-7 cells were co-transfected with CD8-SLAM3 construct and SAP-deletion mutant constructs del1, del2, del3 or del4 (Fig. 1). Cell-surface proteins were biotinylated and cells were lysed 48 h after transfection. Postnuclear lysates were immunoprecipitated with 1 μ g anti-CD8 antibody (OKT8) and immunoprecipitates were immunoblotted with anti-Flag antibody (Kodak) (1/1,000) (top panel) or streptavidin (bottom panel).

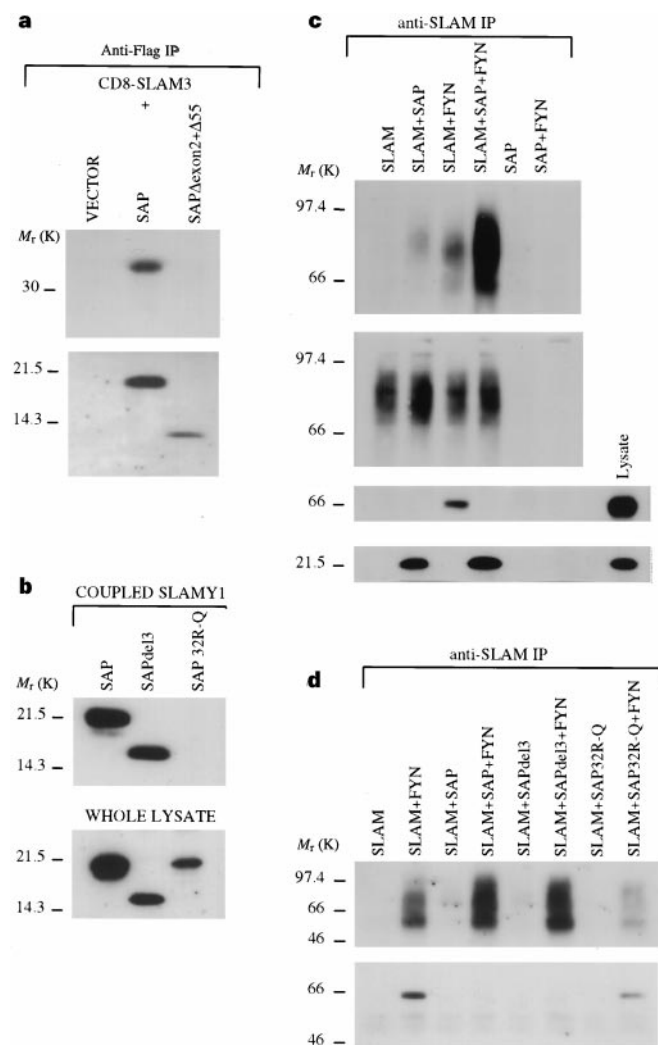


Figure 8 SAP blocks recruitment of the tyrosine phosphatase SHP-2 to the phosphorylated cytoplasmic domain of SLAM. COS-7 cells were transfected as indicated, cell-surface biotinylated and then lysed. Postnuclear lysates were **a**, immunoprecipitated with anti-Flag to isolate SAP and the associated CD8-SLAM3 chimera; immunoprecipitates were immunoblotted with either streptavidin (upper panel) or anti-Flag (lower panel); or **b**, incubated for 1 h with the SLAM Y1 peptide coupled to beads (7 μM). Bead-associated proteins were extracted and immunoblotted with anti-Flag (1/1,000). **c**, **d**, COS-7 cells were co-transfected with a combination of SLAM, SAP and *c-fyn* constructs. 48 h after transfection, cell-surface-expressed proteins were biotinylated, and cells were lysed as described in Methods. Postnuclear lysates were immunoprecipitated with anti-SLAM antibodies and immunoprecipitates were immunoblotted with anti-phosphotyrosine-HRP (Zymed), streptavidin-HRP (Zymed), rabbit anti-SHP2 (Santa Cruz) or rabbit sera anti-human SAP. Note that SLAM is seen here as a broad band because of its extensive glycosylation¹.

shown). Thus, recruitment of SHP-2 to phosphorylated SLAM was blocked by binding of SAP to the sequence surrounding the most membrane-proximal tyrosine residue, Y281.

As SHP-2 can act as a negative regulator of signal transduction¹⁵, binding of SAP could have a positive effect on co-stimulation by anti-SLAM antibody. To test this, we overexpressed SAP by transient transfection into the Jurkat T cells, together with an interleukin (IL)-2-promoter/luciferase reporter construct and SLAM. Stimulation of these transfected cells with a combination of anti-SLAM and anti-CD3 monoclonal antibodies significantly increased IL-2-promoter/luciferase reporter activation compared with Jurkat cells transfected with SLAM alone (Fig. 9). Anti-SLAM by itself did

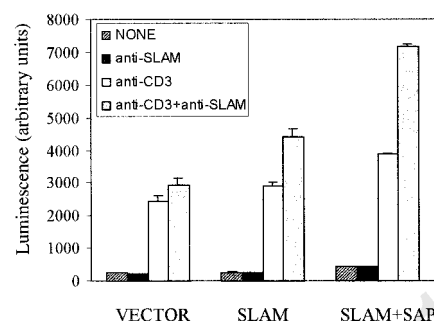


Figure 9 SAP has a positive effect in the SLAM co-stimulatory pathway. Jurkat cells were transfected with SLAM, SAP and SLAM+SAP constructs or vector (pCDNA3), as described in Methods. Luciferase activity was measured after stimulation with the indicated antibodies.

not induce the IL-2-promoter/luciferase reporter signal, and the signal induced by anti-CD3 reagents was weaker than that generated by the two antibodies together. Thus, overexpression of SAP in T cells has a positive effect on the co-stimulatory activity of anti-SLAM.

Discussion

We have identified the gene that is altered in XLP as SAP by using an indirect, functional approach. We have shown that SLAM is the cell-surface protein to which the XLP gene product binds and discovered a new T-cell signal-transduction pathway initiated by the co-receptor molecule SLAM. Binding of the T-cell protein SAP to the cytoplasmic domain of SLAM blocks recruitment of SHP-2. Therefore two modes of SLAM signalling are likely to exist: one in which the inhibitor SAP acts as a negative regulator and another in which SHP-2-dependent signal transduction operates. A switch between these two might occur upon T-cell activation, because expression of the SAP gene drops rapidly after the T-cell antigen receptor is stimulated (C.W., J.S. and C.T., manuscript in preparation). It is possible that SAP is released from SLAM upon activation of T cells to bind to other protein(s). The inhibition by SAP of SH2-domain interactions could be generally extended to molecules other than SLAM. Our results support a model in which SAP controls signal-transduction pathways that are initiated by interactions between SLAM molecules on the interface between T and B cells.

The failure of the immune system to eliminate EBV-infected B cells in XLP appears not to be caused by a B-cell-specific defect, but rather by defective EBV-specific helper-T-cell and cytotoxic-T-cell responses^{10-12,16}. As SAP is found predominantly in thymus-derived lymphocytes, SAP mutations should affect SLAM-induced signal-transduction events in T lymphocytes. Many XLP patients have impaired interferon-γ production by helper T cells, indicative of a Th2-like phenotype. Because engagement of SLAM during antigen-specific T-cell stimulation induces IFN-γ production and redirects the Th2 phenotype to a Th1/Th0 phenotype⁴, an inadequate response of the T-helper subset in XLP patients could result from impaired functioning of the SLAM/SAP pathway. Alternatively, as reduced EBV-specific cytotoxic-T-cell responses have been reported in XLP patients, EBV-infected B cells may not be 'licensed' to activate T-killer cells by helper T cells that lack SAP¹⁷⁻¹⁹. In that model, SLAM-SLAM binding in T/B lymphocyte interactions might serve the same purpose as CD40-CD40L interactions in activation of dendritic cells by helper T cells. On the basis of our results, particularly the correlation between the XLP phenotype and loss of binding of the respective SAP mutants to the SAP docking site on SLAM, we conclude that dysfunctional SLAM/SAP-induced signal-transduction pathways are responsible for the ineffective T-cell response in sustaining elimination of EBV-infected B cells.

Uncontrolled B-cell proliferation in XLP patients frequently leads

to non-Hodgkin's lymphomas induced by EBV. This type of non-Hodgkin's lymphoma is often found in AIDS patients, bone-marrow-transplant patients, and individuals who undergo suppressive T-cell treatment. Some of these disease states could be caused by abnormal regulation of the SAP gene and/or of the T-cell subsets that express SAP. □

Methods

Two-hybrid screening. A cDNA encoding the human SLAM cytoplasmic domain was cloned in the pGBT9 vector and used to screen a cDNA library derived from the human T-cell line KT3 in the yeast two-hybrid system²⁰. The 5' end of human SAP was obtained by rapid amplification of closed ends (RACE; Clontech). Mouse SAP cDNA was cloned by screening a mouse thymus cDNA library in vector Zap Express (gift from L. Clayton) using as a probe a cDNA amplified from mouse thymic mRNA by RT-PCR, obtained with specific primers flanking the 5' and 3' ends of the expressed-sequence tag (EST) AA255258.

RNA and DNA analysis. For northern blotting, poly(A)⁺ RNA from various human tissues (Clontech) or from total mRNA from cell lines (Trizol, GIBCO-BRL) was used as described²¹ with a radiolabelled human SAP-specific cDNA probe. RT-PCR was done with a GenAmp RT-PCR kit (Perkin Elmer) by using the following primers: forward 5'-GCC TGG CTG CAG TAG CAG CGG CAT CTC CC-3'; reverse, 5'-ATG TAC AAA AGT CCA TTT CAG CTT TGA C-3'. The RT-PCR products were then electrophoresed on 10% polyacrylamide gels at 200 volts for 7 h. Gels were stained with CYBR green (Molecular Probes) for 20 min according to the manufacturer's instructions and visualized and photographed under a UV illuminator. For exon PCR and restriction analysis, each of the four exons of the human SAP gene was amplified by PCR using the following primers: exon 1, forward, 5'-GCC CTA CGT AGT GGG TCC ACA TAC CAA CAG-3', and reverse 5'-GCA GGA GGC CCA GGG AAT GAA ATC CCC AGC-3'; exon 2, forward, 5'-GGA AAC TGT GGT TGG GCA GAT ACA ATA TGG-3', and reverse, 5'-GGC TAA ACA GGA CTG GGA CCA AAA TTC TC-3'; exon 3, forward, 5'-GCT CCT CTT GCA GGG AAA TTC AGC CAA CC-3', and reverse, 5'-GCT ACC TCT CAT TTG ACT TGC TGG CTA CAT C-3'; exon 4, forward, 5'-GAC AGG GAC CTA GGC TCA GGC ATA AAC TGA C-3', and reverse, 5'-ATG TAC AAA AGT CCA TTT CAG CTT TGA C-3'. The design of these PCR primers was based on the genomic sequence of human SAP (EMBL database). The PCR products were visualized on 2% agarose gels, subcloned into pCR2.1 vector (Invitrogen), and sequenced. For analysis of a point mutation in the region preceding exon 2, exon 2 PCR products were gel-purified (Qiagen) and digested with the restriction enzyme *MnII* (New England Biolabs), followed by 10% polyacrylamide gel electrophoresis at 200 volts for 3 h.

Rabbit antisera. SLAM Y2, SLAM Y3, hSAP (CQGTGIREDPDV) and mSAP (CQAPTGRDSDI) peptides were coupled to KLH (PIERCE) and used to inject New Zealand rabbits. Rabbit antisera were obtained by standard procedures.

Plasmid construction, transfection and luciferase assays. Human SAP cDNA was cloned in vector pGEX2T (Pharmacia) to generate a GST-SAP construct. Human SAP cDNA and mutant SAP cDNAs, generated by PCR (SAPdel1, SAPdel2, SAPdel3, SAPdel4), were cloned in expression vector pCMV2-Flag (Kodak). CD8-SLAM constructs were generated by PCR as described²² and cloned in pCDNA3 (Invitrogen). The SLAM construct was generated by subcloning human SLAM cDNA in vector pJFE14-SRα (a gift from DNAX Research Institute). E1-4 cells were stably transfected by electroporation with a cDNA encoding SLAM4 or vector alone. Cells were then selected by growth in medium with 600 mg ml⁻¹ neomycin for three weeks, stained with anti-SLAM phycoerythrin-conjugated antibodies (A12), and the positive cells sorted. COS-7 cells were transfected by the DEAE-dextran method²²; cDNAs coding for SLAM and SAP were transiently transfected with an IL-2 promoter-luciferase construct into Jurkat-Tag cells as described²³. After 24 h, transfected cells were stimulated with anti-CD3 and/or anti-SLAM monoclonal antibodies as described¹. Postnuclear lysates were analysed in a Berthold luminometer 8 h later.

Peptide-affinity columns. Whole lysates from untransfected EL-4 cells and from SLAM-transfected EL4-cells were incubated for 3 h with 5 µg GST or GST-SAP protein in the presence of glutathione beads, which were washed

three times and separated in SDS-PAGE as described below. Peptides SLAM Y1 (CVEKKSILTYAQVQK), SLAM pY1 (CVEKKSILTYAQVQK), SLAM F1 (CVEKKSILTYAQVQK), SLAM Y2 (CTTIYVAATEPVPEVQK) and SLAM Y3 (CTVYASVTLPEK) were synthesized and coupled to the beads (sulpholink-coupling gel; Pierce). Coupled peptides were incubated for 1 h with lysates from mouse thymocytes or from CD8-SLAM3 transiently transfected COS-7 cells, in the presence of different concentrations of soluble peptides as indicated. Beads were washed three times and separated by SDS-PAGE.

Immunoprecipitation. Cell lysates were prepared as described²², clarified by centrifugation at 14,000g for 15 min at 4 °C, and the crude lysate was then precleared using 50 µl protein G-agarose beads (GIBCO-BRL) and 5 µl normal mouse serum or normal rabbit serum for 1 h. Immunoprecipitations were done by using the indicated antibodies and 30 ml protein G-agarose beads for 3 h at 4 °C. Beads were then washed as described²². Crude lysates and immunoprecipitates were subjected to SDS-PAGE and transferred onto PVDF filters (Millipore). Filters were blocked for 1 h with 5% skim milk (or 3% bovine serum albumin) and then probed with the indicated antibodies. Bound antibody was revealed using horseradish peroxidase-conjugated secondary antibodies using enhanced chemiluminescence (Supersignal; Pierce). For anti-phosphotyrosine blotting, we used a directly conjugated horseradish peroxidase antibody cocktail (Pierce).

Received 3 July; accepted 8 September 1998.

- Cocks, B. G. *et al.* A novel receptor involved in T-cell activation. *Nature* **376**, 260–263 (1995).
- Aversa, G. *et al.* SLAM and its role in T cell activation and Th cell responses. *Immunol. Cell Biol.* **75**, 202–205 (1997).
- Aversa, G., Chang, C. J., Carballido, J. M., Cocks, B. G. & de Vries, J. E. Engagement of the signalling lymphocytic activation molecule (SLAM) on activated T cells results in IL-2-independent, cyclosporin A-sensitive T cell proliferation and IFN-γ production. *J. Immunol.* **158**, 4036–4044 (1997).
- Carballido, J. M. *et al.* Reversal of human allergic T helper 2 responses by engagement of signaling lymphocytic activation molecule. *J. Immunol.* **159**, 4316–4321 (1997).
- Isomaki, P. *et al.* Increased expression of signaling lymphocytic activation molecule in patients with rheumatoid arthritis and its role in the regulation of cytokine production in rheumatoid synovium. *J. Immunol.* **159**, 2986–2993 (1997).
- Punnonen, J. *et al.* Soluble and membrane-bound forms of signaling lymphocytic activation molecule (SLAM) induce proliferation and Ig synthesis by activated human B lymphocytes. *J. Exp. Med.* **185**, 993–1004 (1997).
- Ferrante, P. *et al.* Cytokine production and surface marker expression in acute and stable multiple sclerosis: altered IL-12 production and augmented signaling lymphocytic activation molecule (SLAM)-expressing lymphocytes in acute multiple sclerosis. *J. Immunol.* **160**, 1514–1521 (1998).
- Lamartine, J. *et al.* Physical map and cosmid contig encompassing a new interstitial deletion of the X-linked lymphoproliferative syndrome region. *Eur. J. Hum. Genet.* **4**, 342–351 (1996).
- Porta, G. 4.5-Mb YAC STS contig 50-kb resolution, spanning Xq25 deletions in two patients with lymphoproliferative syndrome. *Genome Res.* **7**, 27–36 (1997).
- Lanyi, A. *et al.* A yeast artificial chromosome (YAC) contig encompassing the critical region of the X-linked lymphoproliferative disease (XLP) locus. *Genomics* **39**, 55–56 (1997).
- Purtilo, D. T. Immunopathology of X-linked lymphoproliferative syndrome. *Immunol. Today* **4**, 291–295 (1983).
- Seemayer, T. A. *et al.* X-linked lymphoproliferative disease: twenty-five years after the discovery. *Res.* **38**, 471–478 (1995).
- Rousset, F., Souillet, G., Roncarolo, M. G. & Lamelin, J.-P. Studies of EBV-lymphoid cell interactions with the X-linked lymphoproliferative syndrome: normal EBV-specific HLA-restricted cytotoxicity. *Clin. Exp. Immunol.* **63**, 280–289 (1986).
- Cooper, D. M. & Krawczak, M. *Human Gene Mutations* (Bios Scientific, Oxford, 1993).
- Marengere, L. E. *et al.* Regulation of T cell receptor signaling by tyrosine phosphatase SYP association with CTLA-4. *Science* **272**, 1170–1173 (1996).
- Sutkowski, N. *et al.* An Epstein-Barr virus-associated superantigen. *J. Exp. Med.* **184**, 971–980 (1996).
- Ridge, J. P., Di Rosa, F. & Matzinger, P. A conditioned dendritic cell can be a temporal bridge between a CD8⁺ T-helper and a T-killer cell. *Nature* **393**, 474–478 (1998).
- Schoenberger, S. P., Toes, R. E. M., van der Voort, E. I. H., Oeffinger, R. & Melief, C. J. M. T-cell help for cytotoxic T lymphocytes is mediated by CD40–CD40L interactions. *Nature* **393**, 480–483 (1998).
- Bennett, S. R. M. *et al.* Help for cytotoxic T-cell responses is mediated by CD40 signalling. *Nature* **393**, 478–480 (1998).
- Ramesh, N., Anton, I. M., Hartwig, J. H. & Geha, R. S. WIP, a protein associated with Wiskott-Aldrich syndrome protein, induces actin polymerization and redistribution in lymphoid cells. *Proc. Natl Acad. Sci. USA* **94**, 14671–14676 (1997).
- Wang, B. *et al.* Over-expression of CD3-ε transgenes blocks T lymphocyte development. *Int. Immunol.* **7**, 435–448 (1995).
- de Aoz, I. *et al.* Tyrosine phosphorylation of the CD3-ε subunit of the T cell antigen receptor mediates enhanced association with phosphatidylinositol 3-kinase in Jurkat cells. *J. Biol. Chem.* **273**, 25310–25318 (1998).
- Martinez-Martinez, S. *et al.* Blockade of T-cell activation by dithiocarbamates involves novel mechanisms of inhibition of nuclear factor of activated T cells. *Mol. Cell. Biol.* **17**, 6437–6447 (1997).

Acknowledgements. We thank J. David, J. Fingerhuth and N. Letvin for critically reviewing the manuscript; J. Sumegi for sharing unpublished data; N. van Orsouw and J. Vijg for guidance in analysing PCR products; and E. Wang, M. Comiskey, A. Bywater and A. Hren for technical assistance. This work was supported by a grant from the NIH. J.S. is supported by a fellowship from the Ministerio de Educacion y Cultura de España.

Correspondence and requests for materials should be addressed to J.S. (e-mail: jortega@bidmc.harvard.edu) or C.T. (e-mail: terhorst@bidmc.harvard.edu). The sequences reported here will appear in the Genbank nucleotide sequence database under accession numbers AF072930, AF073019 (human SAP) and AF072903 (murine SAP).

The atomic structure of the bluetongue virus core

Jonathan M. Grimes^{*†}, J. Nicholas Burroughs[‡], Patrice Gouet^{**†}, Jonathan M. Diprose^{*}, Robyn Malby^{*}, Stephen H. Zientara[§], Peter P. C. Mertens[‡] & David I. Stuart^{*||}

^{*} The Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

[‡] Institute of Animal Health, Pirbright Laboratory, Ash Road, Pirbright, Woking GU24 0NF, UK

[§] C.N.E.V.A. Alfort, Laboratoire Central de Recherches Vétérinaires, 22 rue Pierre Curie, BP67, 94703 Maisons Alfort Cedex, France

^{||} Oxford Centre for Molecular Sciences, New Chemistry Building, South Parks Road, Oxford OX1 3QT, UK

The structure of the core particle of bluetongue virus has been determined by X-ray crystallography at a resolution approaching 3.5 Å. This transcriptionally active compartment, 700 Å in diameter, represents the largest molecular structure determined in such detail. The atomic structure indicates how approximately 1,000 protein components self-assemble, using both the classical mechanism of quasi-equivalent contacts, which are achieved through triangulation, and a different method, which we term geometrical quasi-equivalence.

Bluetongue virus (BTV) is the prototype species of the genus *Orbivirus* within the Reoviridae, one of the largest families of viruses¹. Members of the family Reoviridae include agents of economically important disease in both plants and animals, and are responsible for significant levels of child mortality in developing countries. BTV infects mammalian hosts (sheep, for example) and is transmitted by insect vectors (*Culicoides* species). Viruses of the Reoviridae are characterized primarily by their genome of 10–12 segments of linear, double-stranded (ds) RNA. Almost all of these separate segments represent single genes, generating a total of 10–13 viral proteins. The Reoviridae are non-lipid-containing icosahedral viruses, usually with an outer capsid layer and a core that contains the genome, which comprises a single copy of each dsRNA segment². The outer capsid has a primary role in cell attachment and penetration. It is relatively variable within and between the different genera of the Reoviridae, reflecting its interactions with different target organs and cell types of widely different host species. However, core particles of BTV from which the outer capsid has been removed, although they show greatly reduced infectivity for mammalian cells, retain an unexpectedly high infectivity for the insect vector³, indicating that core proteins can also mediate cell attachment and penetration.

Cell entry involves outer capsid removal, leading to release of the core into the cytoplasm of the target cell. Because the dsRNA would trigger host defence mechanisms if released into the cytoplasm, dsRNA viruses retain their genomes within a protein core that contains active dsRNA-dependent single-stranded (ss) RNA polymerases (transcriptases) and (for most eukaryotic dsRNA viruses) capping enzymes⁴. The core particle is therefore vital for function, acting as a molecular engine that produces full-length, capped, messenger RNA copies simultaneously from each of the genome segments it contains. The extruded mRNAs function not only as templates for viral protein synthesis, but are also encapsidated within nascent cores, where they act as templates for negative-strand synthesis, producing the progeny dsRNA genome segments. At present, we do not understand how one copy of each gene segment is packaged per particle. The BTV core has a diameter of 700 Å and is composed of two principal structural proteins^{5,6}.

To facilitate protein identification and wider comparisons between viruses, we use extended protein names, appending a tag denoting their structural or functional role⁷. Each particle contains

780 copies of the VP7(T13) protein, of relative molecular mass 38,000 (M_r 38K), arranged as trimers on a $T = 13$ quasi-equivalent lattice^{8–10}, which form the ‘bristly’ core surface. This outer VP7(T13) layer cloaks a thin, rather featureless ‘subcore’ shell^{9,11,12} constructed from 120 copies¹³ of the VP3(T2) protein (M_r 100K). This VP3(T2) subcore can retain its structure if the outer VP7(T13) core layer is lost^{11,12}. In isolation, VP7(T13) can form flat crystalline arrays¹⁴, but VP3(T2) will self-assemble to form particles (ref. 15, and P. Roy, personal communication). In addition, the VP3(T2) molecules bind RNA¹², and the smallest viral particles containing RNA are VP3(T2) subcores¹⁶. The VP3(T2) molecule therefore seems to play a fundamental role in the early stages of formation of the BTV core. The core also contains ‘minor’ enzymatically active proteins and ten segments of dsRNA (~19,000 base pairs (bp) in total).

The BTV core presents the puzzle of how a large complex, containing mismatched symmetry, can assemble. The icosahedral building block is composed of 13 subunits of VP7(T13) attached to two molecules of VP3(T2), such that there must be 13 distinct sets of interactions between VP7(T13) and VP3(T2) in the assembled core. A further problem common to dsRNA viruses is how 120 subunits of VP3(T2) build an icosahedral shell, an ability that is inexplicable within the current conceptual framework¹⁰. It is also unclear whether the VP7(T13) layer is assembled using the conformational switching mechanism seen in other viruses^{17,18}. The analysis reported here solves these puzzles.

Structure determination

The determination of the structure of BTV 1(SA), using X-ray crystallography, is briefly described in the Methods section. Data collection was greatly helped by the availability of an undulator beamline (ID-2) at the European Synchrotron Radiation Facility. By combining information from a 22-Å resolution cryo-electron-micrographic reconstruction of BTV 10(USA)⁹ and a high-resolution crystal structure for the VP7(T13) trimer¹⁹, a model was constructed that provided phase information to far higher resolution than the electron-micrographic reconstruction. This provided a route to an atomic model without the computational labour of real-space phase extension. At present, the structure is partly refined at approaching 3.5 Å resolution (see Methods). The current atomic model consists of all residues from the 13 subunits of VP7(T13) and residues 57–901 and 1–803, 814–901 of the A and B molecules, respectively, of VP3(T2) (referred to below as VP3(T2)A and VP3(T2)B; Fig. 1), which together make up the icosahedral asymmetric unit that totals 49,062 non-hydrogen

[†] Present addresses: EMBL Grenoble Outstation, c/o ILL, BP156, 38042 Grenoble, Cedex 9, France (J.M.G.); C.N.R.S.-I.P.B.S., Groupe de cristallographie biologique, 205 route de Narbonne, 31077 Toulouse Cedex, France (P.G.).

atoms. At lower resolution, the electron density reveals disordered protein and RNA features on the inside of the VP3(T2) layer. These features will be presented elsewhere.

Overall architecture of the BTV core

As expected, the core capsid obeys icosahedral symmetry. The overall features of the structure, a lattice of VP7(T13) trimers coating a thin inner shell of VP3(T2) monomers, enclosing the genome, are shown in Fig. 1. The inner surface of the VP3(T2) layer is relatively bland with some shallow, sculpted grooves and a few charged residues. The core possesses small pores (diameter 7 Å) at the icosahedral three-fold axes and larger openings (9 Å diameter) at the five-fold axes. A simple rearrangement of the side chain of Arg154(VP3(T2)B) would enlarge the three-fold pores to some 15 Å. Although these might act as portals in the subcore, the T trimers of VP7(T13) tightly plug these holes in the complete core (Fig. 1). In contrast, whereas the pores at the five-fold axes are unencumbered by molecules of VP7(T13), they are too small to allow exit of nascent mRNA and cannot be significantly enlarged by side-chain movements alone. Activation of the cores for transcription is therefore likely to be accompanied by conformational changes at the five-fold axes, to open these pores further. This agrees with the crystallographic mobility parameters, which show that the core is particularly flexible around the five-fold axes, with the VP7(T13) trimers riding upon the VP3(T2) layer and acquiring further flexibility in the exposed upper domains (Fig. 1c). These pores are lined with four arginines from each five-fold-related VP3(T2)A, of which three are strictly conserved across the orbiviruses (BTV residues 308, 317 and 413), whereas the other (431) is found in BTV, African horse sickness virus (AHSV) and epizootic haemorrhagic disease virus (EHDV) (Fig. 2a, b). These positively charged groups may steer the RNA electrostatically, and may be important in stripping counter ions from any RNA entering or exiting the particle. In line with this we find that the particle is barely permeable to Cs⁺ ions (unpublished data).

The structure of the VP3(T2) molecule

The VP3(T2) subunits are arranged with icosahedral symmetry within the subcore layer, forming two sets (A and B) of 60 subunits each. Both the VP3(T2)A and VP3(T2)B subunits have a similar

overall structure, resembling a triangular wedge of approximately 130 Å by 75 Å, with a thickness varying between 13 Å at the icosahedral two-fold and three-fold axes and 35 Å close to the five-fold axes. The 901 amino acids of VP3(T2) are organized into a unique plate-like structure of three domains—apical (close to the five-fold axes of the icosahedron), carapace (forming a rigid plate) and dimerization (forming the new quasi-two-fold interaction that will be discussed below)—stacked along the long axis of the triangle. The course of the polypeptide chain reveals the apical domain to be an insert in the amino-terminal half of the carapace and, conversely, the dimerization domain to be an insert in the carboxy-terminal half (Fig. 2c, d). The three domains have quite different folds from each other and appear to be unlike any other reported structures²⁰. VP3(T2) is rich in 3_{10} helices, containing 17 compared with 54 α -helices in total between the two copies, A and B. For simplicity, the terms helix and α are used generically below to include both types.

The apical domain (residues 298–587) is a mixture of β -strands and helices (molecule A, 10 β -strands plus 11 helices; molecule B, 11 β -strands plus 10 helices). This domain is particularly rich in residues that are conserved across the orbiviruses (Fig. 2b), and yet it is the site of the most substantial conformational differences between VP3(T2)A and VP3(T2)B, as discussed below. Furthermore, the domain has a complex internal structure and includes the residues that form the putative mRNA exit pore, all of which suggest that the domain carries a considerable functional load. The N-terminal 56 residues of VP3(T2)A are not clearly resolved in the electron density map. Diffuse density indicates that they have switched position and conformation from the carapace domain to the inside of the apical domain, lying adjacent to the five-fold axes.

The carapace (residues 7–297, 588–698 and 855–901) is extensive, covering an area of roughly 50 Å by 70 Å, and yet combines thinness with rigidity (this domain is the least perturbed between molecules VP3(T2)A and VP3(T2)B). This is achieved by forming a simple two-layer structure composed mainly of helices, 20 in A and 21 in B, arranged in a mat-like structure, with the 'warp' and 'weft' cross-braced although not intertwined. Some of the helices are long (for instance, α -15c contains 28 residues; Fig. 2a, c) and stretch across the entire width of the molecule. In addition to the helical core of the domain, the base, which abuts the dimerization domain,

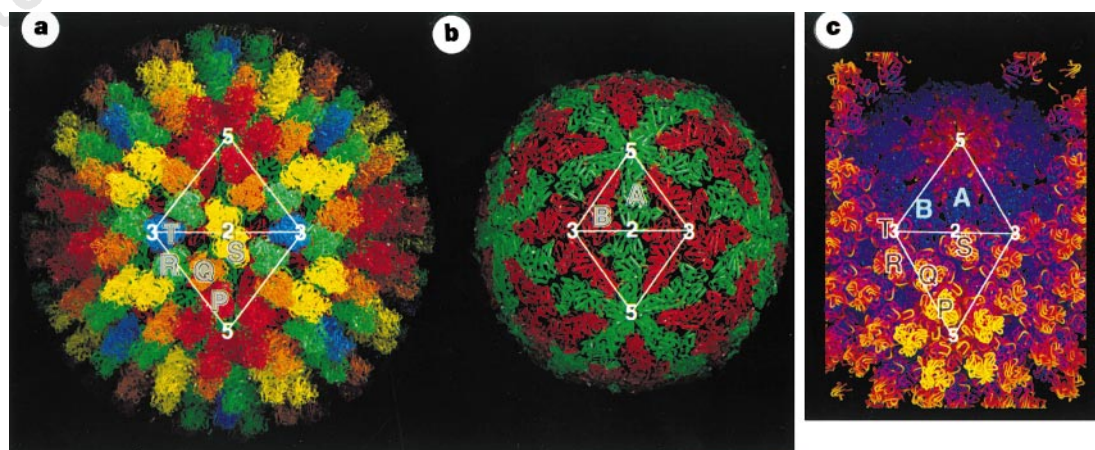


Figure 1 Essential features of the architecture of the BTV core. **a**, The core particle of BTV 1. The icosahedral asymmetric unit (which is a triangular area defined as marked by the symmetry axes of the icosahedron) contains 13 copies of VP7(T13) arranged as 5 trimers, P, Q, R, S and T, coloured red, orange, green, yellow and blue, respectively. Trimer T sits on the icosahedral three-fold axis and thus contributes a monomer to the unique portion. The inner layer of VP3(T2) molecules is coloured red and green. Helices are shown as rods and β -strands as

arrows. **b**, The VP3(T2) scaffold. The icosahedrally unique molecules A and B are coloured in green and red, respectively. Note the completely different structural environment of the A and B molecules. **c**, Cut-away diagram that is colour-coded by *B*-factor. Blue indicates a low *B*-factor (<10 Å²), yellow indicates a high *B*-factor (>150 Å²). VP3(T2) is displayed as a ribbon and VP7(T13) as a thread. Note the increase in mobility of both the VP3(T2) and VP7(T13) layers closer to the icosahedral five-fold axes and for the upper domains of VP7(T13).

is rich in β -structure (18 strands in both A and B) and forms a β -sandwich with an adjoining β -sheet from the dimerization domain. Three of the four chain termini for the VP3(T2)A and VP3(T2)B molecules are contained within this domain and lie internal to the capsid.

The dimerization domain (residues 699–854) contains five helices in A and four in B, but the heart of the domain is a pair of β -sheets, 13 strands in A and 14 in B, disposed perpendicular to each other and joined by one strand that twists between the two. The larger β -sheet, farther away from the dimer interface, packs against the β -sheet that spans the lower portion of the carapace domain, forming a β -sandwich (Fig. 2c).

T = 2 geometrical quasi-equivalence

The arrangement of the 120 VP3(T2) molecules on an icosahedrally symmetric lattice is shown in Fig. 1b. Pairs of A subunits span icosahedral two-fold axes (where they almost touch), linking adjacent five-fold axes, to form an almost continuous scaffold that defines the size of the subcore (Fig. 1b). The B subunits form triangular plugs at the icosahedral three-fold axes, sealing off the subcore (Fig. 1b). This architecture may be unique to dsRNA viruses, among which it seems to be highly conserved; for instance, the overall disposition of the 120 copies of the Gag subunits of a virus from a different family of dsRNA viruses, L-A virus²¹, is strikingly similar.

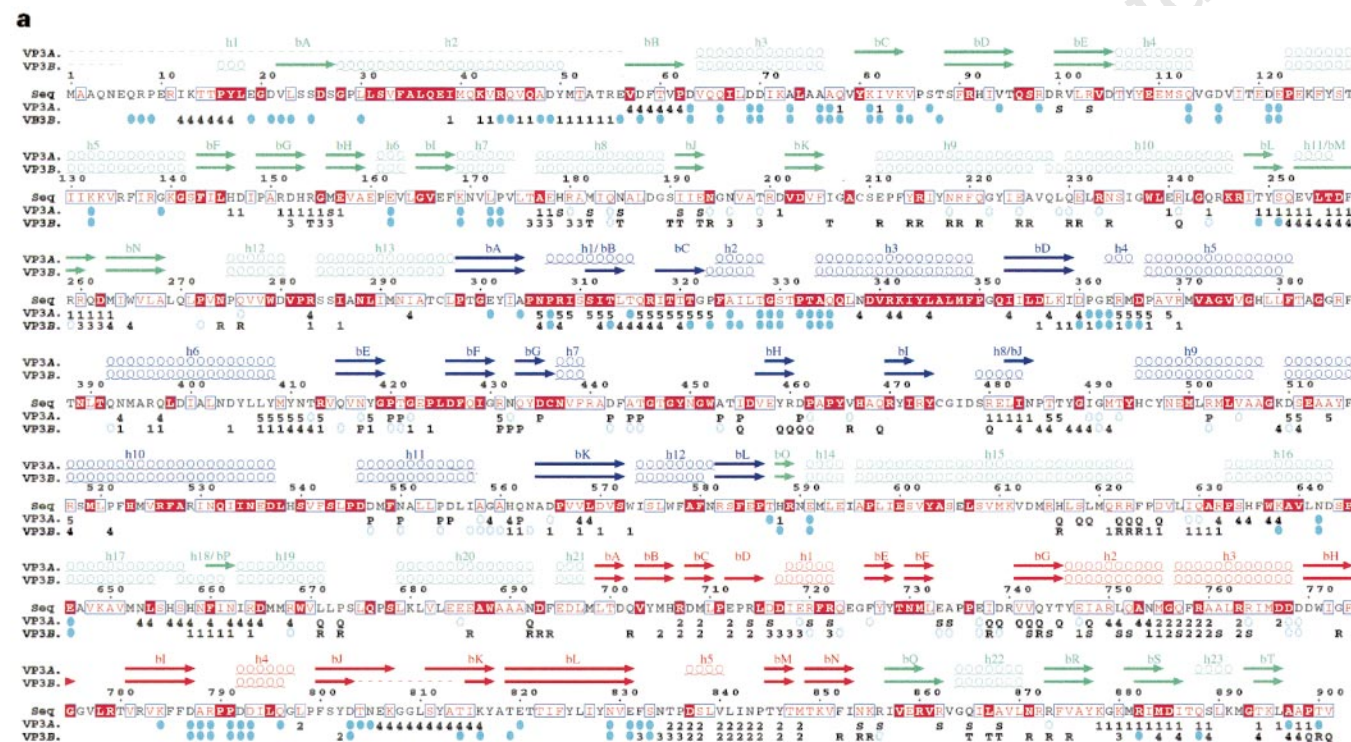


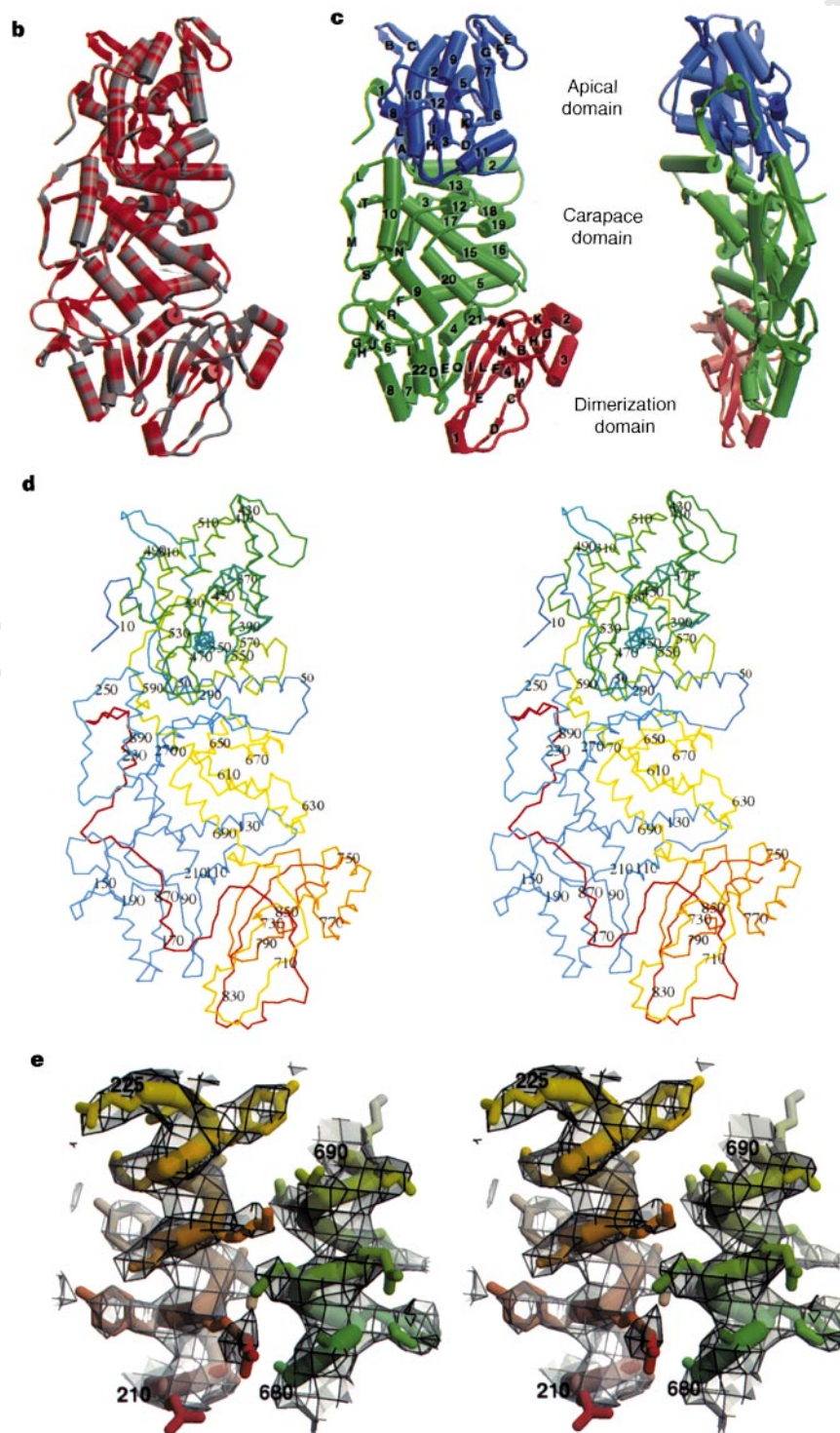
Figure 2 The VP3(T2) molecule. **a**, Sequence/structure analysis of VP3(T2). The BTv 1 sequence was obtained from complementary DNA sequencing, and electron density maps were of sufficient quality so that, by using the map and ten other orbivirus sequences, mistakes in that sequence were corrected. Residues are boxed according to an alignment made with CLUSTAL-W²⁷ against orbivirus sequences from BTv (SwissProt ID VP3_BT10, VP3_BT17, VP3_BT13, VP3_BT11, VP3_BT2, VP3_BT2M, VP3_BT1a), AHSV (VP3_AHSV4), EHDV (VP3_EHDV1) and Broadhaven virus (VP2_BRD). Residues strictly conserved have a red background, residues well conserved according to a Risler matrix³⁸ are indicated by red letters, and the remainder are in black. Properties for VP3(T2)A and B are displayed above and below the sequence. Secondary structural elements, assigned by eye and checked using STRIDE³⁹, are shown above the sequence with VP3(T2)A uppermost. Helices are represented by squiggles, strands by arrows, and regions for which no electron density was seen by dotted lines (three-dimensional structure for molecules A and B start at residues 57 and 7, respectively). Secondary structure elements are coloured by domain: carapace (1–297, 588–698 and 855–901) is green, apical (298–587) is blue and dimerization (699–856) is red. Symbols below the sequence define accessibility and molecular contacts, with A uppermost. Relative accessibility (calculated for each residue by dividing DSSP⁴⁰ values by maximal accessibility according to residue type) is shown by ovals. Residues with a relative accessibility >40% are shown with blue ovals. A filled oval indicates a residue located at the inner surface of

VP3(T2) in the core, an open oval indicates a residue located at the outer surface. Letters indicate contact residues (atom separations of less than 4 Å). A green letter denotes a contact of VP3(T2)A with another VP3(T2) molecule (1, 2, 4 and 5 correspond to A–B, A₂–B, A₅–B and A–A, contacts, respectively, as defined in Fig. 4). A blue letter indicates a contact with VP7(T13) (P, Q, R, S or T as indicated). VP3(T2)B contacts are shown with a red letter for interactions with other VP3(T2) molecules (1, 2, 4 as defined above, plus 3 for B–B₃), whereas contacts with VP7(T13) are shown with blue letters. Figure drawn using ESPript (P.G. and F. Métoz, unpublished program). **b**, A representation of the VP3(T2)B molecule, colour-coded by residue conservation across the orbiviruses. Red indicates a residue that is strongly conserved. Note the strongly conserved β -sheet at the five-fold axis, the conserved helix at the dimerization interface and the general pattern of conserved residues being found where two helices come into close contact. **c**, The secondary structural elements of VP3(T2)B. Left, the standard view; right, an orthogonal view. Colouring is by domain: apical is blue (residues 297–588), carapace is green (residues 7–297, 588–698 and 855–901) and dimerization is red (residues 698–855). Note the radial thinness of the carapace, thickening at the dimerization domain. **d**, Stereo image of the C α trace of VP3(T2)B. Colour is ramped by residue from blue at the N terminus to red at the C terminus, and every 20th residue is labelled. **e**, Stereo image showing a representative portion of the electron density map contoured at an arbitrary level. Residues 210–255 and 680–690 are shown for VP3(T2)B.

The architecture does not fit within the dominant hypothesis for virus assembly. Caspar and Klug¹⁰ put forward a theory that explains the architecture of the icosahedral shells of many viruses. In simple terms, this proposes that the basic triangular icosahedral building block is broken down into a number (denoted T) of equilateral sub-triangles, so that the icosahedral shell is composed of $60 \times T$ chemically identical subunits. They propose that the sub-triangles are sufficiently similar in structure and local environment that the same non-covalent interactions may be used for each of the T distinct copies (the interactions being termed quasi-equivalent). The VP3(T2) layer might be thought of as $T = 2$, although this T number is excluded owing to conformational

restrictions, as indicated in Fig. 3a. We find, however, that a $T = 2$ lattice can be formed if distortion is allowed between the two sub-triangles. In this approximation to traditional triangulation, a more substantial distortion precludes exact matching (we suggest the term geometrical quasi-equivalence for this type of assembly) (Fig. 3a).

The distortion required is a simple function of the width of the subtriangle. This is the solution used by BTV, thus Fig. 3a follows roughly the disposition of the VP3(T2) subunits in the B'V core. The VP3(T2) molecule accommodates perfectly the required distortions, being in the form of a flexible elongated triangle (Fig. 3b). Ten of these narrow subunits are arranged around each five-fold



axis, bound intimately together (various interfaces within the core are analysed in Fig. 4, in terms of contact area and surface complementarity²²) to form a decamer. We postulate that the decamer is an early, transient, assembly intermediate for the virus (the earliest stable assembly intermediate reported is the VP3(T2) subcore containing 12 decamers^{11,16,23}). Decamers are clipped together by weaker interactions of the dimerization domains along their outer edges, each interface consisting of two sets of identical contacts, related by an icosahedral two-fold axis. Each set is composed of two principal interactions; the most important is about the local two-fold axis relating VP3(T2)A and VP3(T2)B from different decamers (Fig. 4). VP3(T2)B subunits from adjacent decamers also form a less substantial association around the icosahedral three-fold axes. The local two-fold interaction is truly quasi-equivalent, but is achieved under the constraints of geometrical quasi-equivalence (described above) and the need to form a closed shell, so that the contacts between the A and B subunits within the decamers are chemically distinct.

The distortion of VP3(T2) required to form the closely packed

shell is focused primarily on the dimerization domain where small conformational rearrangements allow a rotation of $\sim 20^\circ$, around helix α -3, which is involved intimately in the packing interface of the quasi-two-fold between molecules VP3(T2)A and VP3(T2)B. This relatively modest pivoting, combined with a subtle rotation of the apical domain, swings the distal tip of the apical domain ~ 35 Å. Although the apical domains of VP3(T2)A and VP3(T2)B are generally similar (root-mean-square (r.m.s.) deviation of 1.1 Å for 91% of the C α s), substantial differences occur in two loops: 305–329 and 476–494 (Fig. 3b). In molecule VP3(T2)A, loop 305–329 forms a tight annulus at the five-fold axis, whereas in VP3(T2)B it fits under the apical domain of VP3(T2)A. Loop 476–494 is supported in VP3(T2)B by the ordered N terminus, whereas in VP3(T2)A the disordered N terminus allows this loop to collapse. There are also some more modest structural rearrangements at the VP3(T2)A/VP3(T2)B interfaces. The various changes between VP3(T2)A and VP3(T2)B combine to mould the subunits into a close-fitting decamer (Fig. 4), while maintaining the almost exact, local two-fold relationship between molecules in adjacent decamers

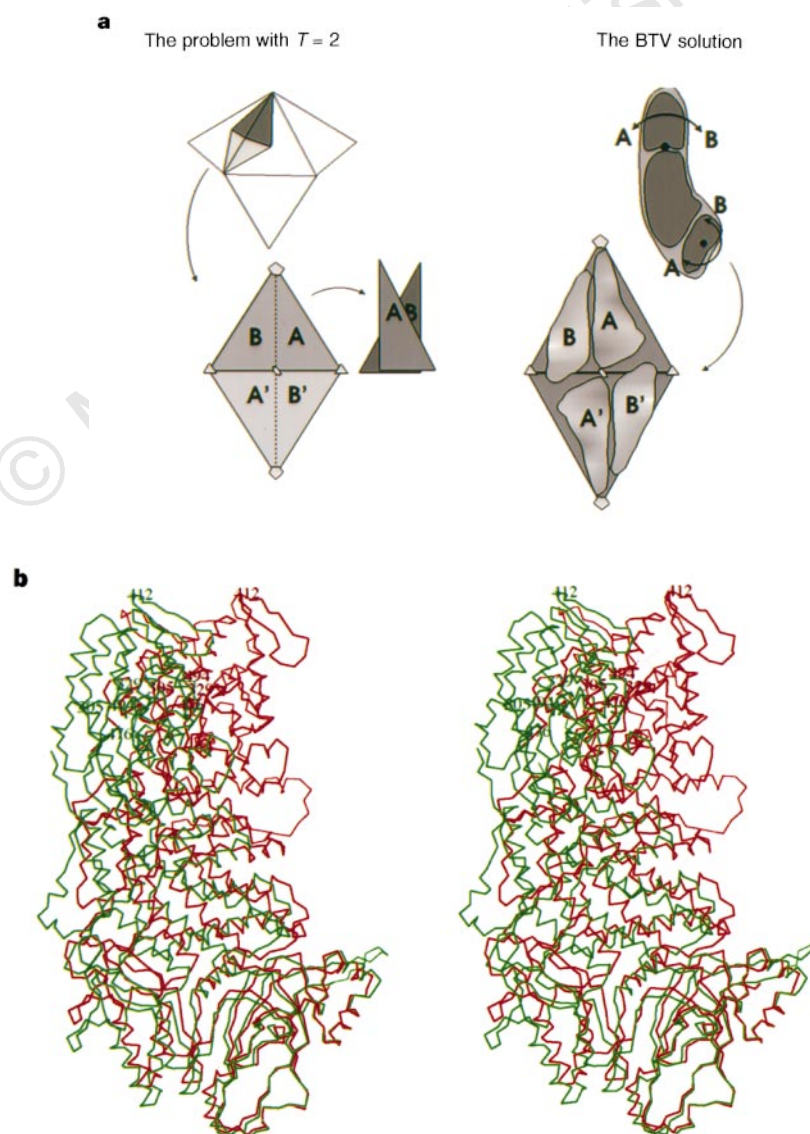


Figure 3 Geometrical quasi-equivalence in the VP3(T2) layer. **a**, Representation demonstrating the problems of applying $T=2$ symmetry using quasi-equivalence. These restrictions may be overcome by a distortion of the molecule, as shown. **b**, Stereo image of the C α chain of VP3(T2), showing the conformational differences between molecules A and B, on superimposition of the structural

elements involved in forming the important quasi two-fold dimerization interface. The traces of molecules A and B are coloured green and red, respectively. Note the 35 Å distortion that has been produced at Thr 412. This consists of both a 31 Å horizontal distortion and a 12 Å difference in radius. In addition, loops 305–329 and 476–494 are particularly different between the two molecules.

that locks the decamers together and is the hallmark of the $T = 2$ geometrical quasi-equivalence.

The VP7(T13) layer

The outer layer of the core is built up of 13 icosahedrally independent copies of the VP7(T13) protein, in the form of four trimers in general positions and one situated with its molecular three-fold axis aligned with the icosahedral three-fold axis (Fig. 1a). These copies are extremely similar to each other (with maximum and minimum r.m.s. deviations in $\text{C}\alpha$ positions between pairs of molecules of 0.30 Å and 0.15 Å, respectively; Fig. 5a), and are almost indistinguishable from the structure observed in monoclinic crystals of the recombinant VP7(T13) protein¹⁹ (r.m.s. deviation 0.37 Å for all $\text{C}\alpha$ s). Thus the $T = 13$ lattice follows the precept of classical quasi-equivalence to an extraordinary degree. The quasi-equivalent trimers, named P, Q, R, S and T⁹ (Fig. 1a), are arranged in pairs about almost exact, local two-fold symmetry axes, which are perpendicular to the surface of the virus particle. The rotations vary from 171°, between P trimers positioned around the five-fold axes, to 180° for the S trimers at the icosahedral two-fold axes. There is a 10 Å upwelling of the VP3(T2) subcore at the five-fold axes, so that not all trimer axes are exactly radial to the particle (Fig. 3b). The maximum deviation is 10° at the P trimer, decreasing smoothly to zero for the trimers located on the icosahedral three-fold axes.

A distortion of the contacts between subunits is inherent in any sub-triangulated icosahedral structure where molecules must adjust to both icosahedral five-fold and local six-fold symmetry axes. This is one reason why viruses whose subunits are arranged with apparently quasi-equivalent geometry do not, in practice, use a single set of atomic interactions to construct the shell. Usually, conformational switching occurs, using alternative stabilizing contacts to provide controlled, path-dependent assembly of the particle^{17,24}. Conformational switching thereby resolves a second conundrum, providing a mechanism by which viruses determine the number of subunits in their capsids. BTV overcomes these two problems by mechanisms that are appropriate to the assembly of a more complex structure. The VP3(T2) layer fulfils the sizing role, setting the number of subunits in the core. The need for different protein conformations in the VP7(T13) layer is eliminated by focusing the interactions between the trimers into a thin band around the larger lower domains (Fig. 5b). This interface is a greasy hinge formed principally by the side-to-side packing of short hydrophobic side chains from helix 4 and its two-fold or quasi

two-fold mate in the neighbouring trimer (Fig. 5c). This facilitates the rolling of trimers against each other, so that the same contacts can be used to form the tilted rings of trimers around the five-fold axes and flat hexagonal rings elsewhere. Residues Ala 89, Gly 92 and Ala 95 are used in 11 of the 13 different trimer interactions, Met 88 and Ile 91 are used 12 times and residue Met 1, which is very close to the symmetry axes relating neighbouring trimers, makes contact in all 13 interactions. The limited contact area of these surfaces (Fig. 5d, legend) indicates that the interactions are relatively flimsy. Nevertheless, isolated trimers appear to have a natural tendency to form hexameric rings that are analogous to those on the virus surface, as these have been seen in *in vivo* crystals of the VP7(T13) protein of AHSV¹⁴ and rice dwarf virus²⁵. The contacts in the virus presumably deviate only slightly from these canonical interactions.

There are three points in the structure where small but significant changes between the 13 copies of VP7(T13) occur, two of which are associated with the packing of the trimers into the core (Fig. 5a). Residues 336–338, on the loop before the C-terminal helix, have high B-factors and adjust slightly to bolster the interactions described above, between adjacent VP7(T13) trimers in the $T = 13$ lattice. Residues 20–26, on a long loop between α -helices 1 and 2 (as defined in ref. 19), are involved in interactions between VP7(T13) trimers (residues 20–22), and they also interact with the underlying VP3(T2) layer, slight adjustments presumably ameliorating the effects of the symmetry mismatch between these two layers. The final site of perturbation, residues 178–179, lies in the ‘neck’ of the VP7(T13) trimer. These changes seem to be caused by a bound moiety, for which weak electron density is seen for all 13 copies of VP7(T13). The density, which also contacts the single essential lysine²⁶ (Lys 255), cannot belong to either VP7(T13) or VP3(T2), but would be consistent with a range of chemical species, for instance a bound oligosaccharide, or even a non-structural protein of the virus. The neck of the VP7(T13) molecule is capable of undergoing conformational changes²⁷, thus the observed density may represent a conformational modulator analogous to the pocket factors observed in picornaviruses²⁸, or it may even correspond to a fragment of bound receptor.

Interactions between VP3(T2) and VP7(T13)

The VP7(T13) and VP3(T2) layers interact through flattish, predominantly hydrophobic, surfaces. The symmetry mismatch means that there are 13 different sets of contacts. Because VP7(T13) subunits trimerize in solution, we may think of the core as being

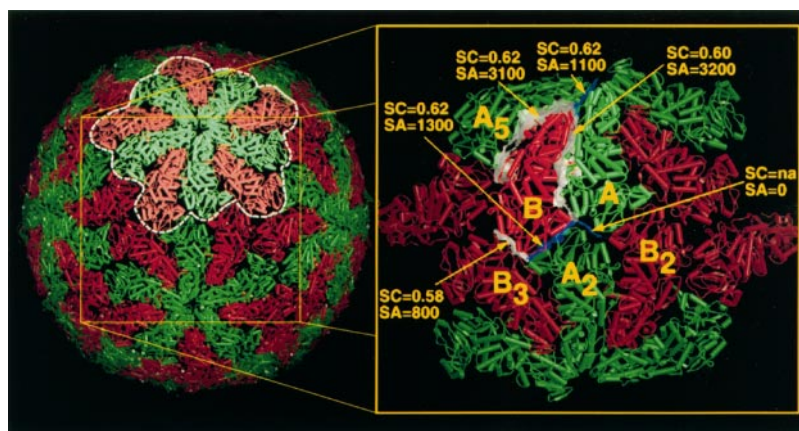


Figure 4 Contacts in the VP3(T2) layer. Surface representation of the VP3(T2) layer highlighting a decamer, possibly an early assembly intermediate. The different VP3(T2)–VP3(T2) contacts have been highlighted in the close-up, and labelled with an assessment of the shape complementarity²² of the interfaces (SC, summed over the atoms involved in the contact with a value of 1 for a perfectly fitting interface and 0 for orthogonal surfaces) and the amount of

surface area⁴⁰ buried in these contacts (SA, Å²) (see Fig. 5d for comparative values for other contact surfaces). This analysis indicates that, on steric considerations alone, the decamer is more likely to form than the A–B₂ dimer, although of course we do not know how much of the binding energy is used in switching the conformation of the A and B molecules from that adopted in isolation.

made from the crystallization of 260 VP7(T13) trimers onto the subcore of 120 VP3(T2) subunits, the driving force being the interaction of each trimer with the underlying VP3(T2) layer. These interactions are displayed schematically in Fig. 5d and quantified in the legend. It is remarkable that the various VP7(T13) trimers remain virtually unperturbed while making completely different, yet extensive and relatively well fitting, contacts with the VP3(T2)A and VP3(T2)B molecules (Fig. 5d). One would expect a hierarchy of interactions, to facilitate the assembly into a properly orientated and closed $T = 13$ lattice. The T trimers seem to be the most tightly attached, and are well placed to act as nucleation points for the binding of VP7(T13) trimers to the VP3(T2) subcore, as the symmetry of the VP3(T2) and VP7(T13) layers match at this point. The size and shape of this subcore then determines the deviations from hexagonal packing, during the ordered addition of further VP7(T13) trimers, to correctly define the $T = 13$ lattice. This agrees with the observation that those trimers (P) most distant from the icosahedral three-fold axes make

the least favourable contacts (Fig. 5d) and indeed are missing from recombinant core-like-particles⁶. This two-dimensional crystallization model avoids the need for the guidance of conformational switching, furnishing an assembly mechanism well suited to the most complex viruses.

Incorporation of minor proteins and RNA

It seems likely that the proper organization of the RNA for efficient transcription requires that it be laid down within a pre-existing protein shell. The limited size of the pores in the assembled subcore would, however, prevent entry of the enzymes. It therefore seems probable that the enzymatic components attach to VP3(T2) decamers, so that upon formation of the complete subcore there is a transcription complex at each of the 12 five-fold axes.

Subcores that are empty of RNA could possess a very different conformation from the final form that we observe. Radial plasticity of the subcore is indicated by the structural variation between the A and B subunits, and agrees with the observation of collapsed

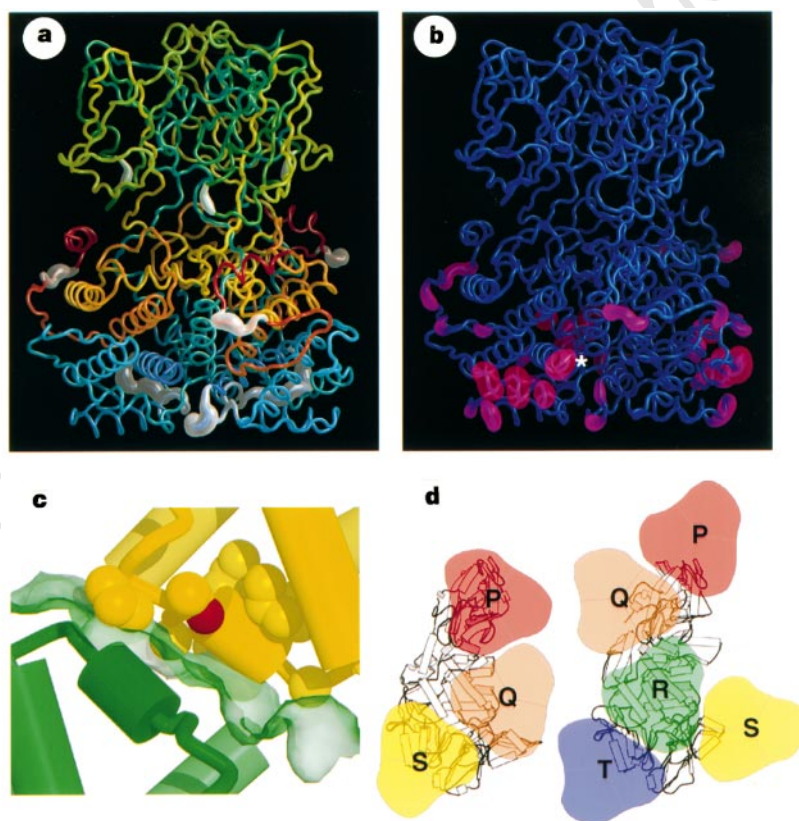


Figure 5 The VP7(T13) layer. **a**, Representation of a trimer of VP7(T13), colour ramped from blue to red by sequence from residues 1 to 349, with the structural differences between all 13 copies of VP7(T13) mapped on as a semi-transparent surface. The width of the coil is the r.m.s. deviation between all 13 independent chains (assuming these deviations to be isotropic) projected onto a radial vector. The maximum deviation of 0.86 Å occurs at residue 21, near the base of the molecule. The other peaks are 0.65 Å at residue 179 (neck) and 0.83 Å at residue 337 (halfway up the base). **b**, Representation of a trimer of VP7(T13), with residues involved in inter-trimer contacts on the core surface indicated by the semi-transparent surface. The width of the coil is proportional to the frequency with which each residue is involved in the 13 different VP7(T13)-VP7(T13) contacts. Residue 1 (marked with an asterisk) is involved in all 13 interactions. **c**, Representation showing the hydrophobic residues involved in forming many of the VP7(T13)-VP7(T13) interactions. These relatively nonspecific contacts allow the VP7(T13) trimers to interact in slightly different ways, enabling them to meet the symmetry requirements for a $T = 13$ layer. One trimer is enveloped in a semi-transparent surface (green) and the crucially important side-chain atoms of the

residues near helix 4 are shown in a CPK rendition in the other trimer. For this second trimer, the representation is shown in yellow, as are carbon atoms of the side chains. Only one oxygen or nitrogen atom is visible, the oxygen of Thr 90. The view is down the two-fold axis relating the trimers. **d**, Representation showing the interactions of VP7(T13) with both other molecules of VP7(T13) and the underlying VP3(T2) layer. SC^{22} and SA (Fig. 4 legend) have been calculated for the various contacts. Values for the various VP7(T13)-VP3(T2) interactions are (SC, SA, ratio of occluded surface on VP3(T2)A: VP3(T2)B): P(0.41, 1,900 Å², 80:20), Q(0.46, 2,500 Å², 60:40), R(0.45, 2,800 Å², 0:100), S(0.45, 2,600 Å², 75:25), T(0.48, 2,400 Å², 0:100). Values for the VP7(T13)-VP7(T13) contacts are (SC, SA): P-P(0.37, 480 Å²), P-Q(0.40, 1,200 Å²), Q-R(0.49, 730 Å²), Q-S(0.51, 470 Å²), R-S(0.53, 540 Å²), R-T(0.56, 800 Å²), S-S(0.35, 1,200 Å²). Interestingly, the values for SC span those seen for two TCR-MHC/peptide complexes (0.45-0.47 (ref. 41)), another weak interaction where the complex can be formed from a range of amino-acid side chains, but are substantially worse than for strong interactions (for example, typically 0.67 for an antibody-antigen complex²²).

Table 1 Structure determination

Total film packs	980	
Total reflections	21,510,811	
Unique reflections	3,299,866	
	Overall (100–3.5 Å)	4.0–3.5 Å
Data completeness	0.535	0.078
R_{merge}	0.229	0.677
Averaging R -factor†	0.160	0.112
Correlation coefficient‡	0.901	0.868

* $R_{\text{merge}} = \sum_i \sum_h |I_{i,h} - \langle I_i \rangle| / \sum_i \sum_h \langle I_i \rangle$, where h are unique reflection indices, $I_{i,h}$ are intensities of symmetry-related reflections and $\langle I_i \rangle$ is the mean intensity³⁰.

† R -factor = $\sum_h |F_o|_h - |F_c|_h| / \sum_h |F_o|_h$, where h are the unique reflection indices, F_o are the observed structure factors and F_c are the structure factors calculated from inversion of the solvent-flattened map³⁰ (SHELLSCALE; D.I.S., unpublished program).

‡Correlation coefficient = $\sum_h (\langle F_o \rangle - |F_o|_h)(\langle F_c \rangle - |F_c|_h) / [\sum_h (\langle F_o \rangle - |F_o|_h)^2 - (\langle F_c \rangle - |F_c|_h)^2]^{1/2}$, where h are the unique reflection indices, F_o are the observed structure factors and F_c are the structure factors calculated from inversion of the solvent-flattened map³⁰ (SHELLSCALE; D.I.S., unpublished program).

structures in the dsRNA phage $\Phi 6$ (ref. 29), which show invagination at the five-fold axes. Structures that are superficially similar to the collapsed pro-capsid of $\Phi 6$ have been observed for orbiviruses¹¹ (P.P.C.M., unpublished data). Such a collapse at the five-fold axes could be accomplished by an amplification of a conformational change that is observed in the core between VP3(T2)A and VP3(T2)B (a bending of the molecule in a radial direction, which in the mature core pushes the apical domain of VP3(T2)A some 10 Å outwards relative to that in VP3(T2)B; Fig. 3b). The increase in flexibility away from the two-fold and three-fold axes (Fig. 1c) also supports such a model.

Conclusions

We have found a solution to the assembly problem of the highly complex BTv core which does not require conformational switching in the VP7(T13) layer. The core is segregated into two layers. A thin inner shell is constructed from 120 large triangular plates of the VP3(T2) protein, with the minimum possible conformational distortion. This represents a generalization of classic assembly by means of triangulation combined with quasi-equivalence, which we term geometrical quasi-equivalence. This fragile internal scaffold carries the information that defines the size of the virus. The core is then made rigid by the adherent lattice of 260 essentially identical trimers of VP7(T13), which assembles by following precisely the principles of quasi-equivalence by epitaxial two-dimensional crystallization. Extensive, but inevitably nonspecific interactions ameliorate the symmetry mismatch between the two layers to produce a robust shell. Such a mixture of assembly modes provides a mechanism appropriate for the assembly of even larger viruses and other macromolecular complexes. □

Methods

Data collection. The growth, purification and crystallization of cores of BTv 1(SA) were done as described previously^{5,13}. Data collection (Table 1) and structure determination will be described in greater detail elsewhere. The crystals (usually between 0.3 and 0.4 mm in the longest dimension) were pre-mounted in quartz capillary tubes at IAH, Pirbright. Data were collected on a Marresearch imaging plate (diameter 30 cm), to low resolution at PX station 9.6, SRS, Daresbury¹³, and subsequently to higher resolution at the undulator beamline ID2, ESRF, Grenoble. Data to a nominal resolution of 3.5 Å Bragg spacings were collected on ID2, by oscillating the crystal through 0.25° at ~8°C. Only one exposure was possible per sample position, owing to radiation damage, and consequently well over 1,000 crystals were examined to obtain these data. The beam cross-section was no more than 0.1 × 0.1 mm² for ESRF data collection. The crystals (space group $P2_12_12$, unit cell dimensions $a = 795.6$ Å, $b = 821.8$ Å and $c = 753.3$ Å), contain half a particle in the asymmetric unit (30-fold non-crystallographic symmetry, n.c.s.).

Map calculation. Data were processed using DENZO³⁰, the resolution determined dynamically with TRIM-DENZO (D.I.S., unpublished computer program) and scaled using an enlarged version of SCALEPACK³⁰. Over

8,000,000 data points were collected, giving a unique data set of some 2,800,000 reflections, which were sharpened (see below). The structure was solved by molecular replacement using as a search model the cryo-electron-micrographic-derived VP7(T13) model of the outer layer of the core⁹, composed of Cα atoms. A self-rotation function (program X-PLOR³¹) provided the orientation of the virus about the z axis. A translation search (X-PLOR) then located the particle in z (correlation coefficient (see Table 1 for definition) of 0.29 for data from 60 to 30 Å at $z = 0.249$). Rigid body refinements of the model, to orientate the molecules and optimize the n.c.s. operators, were performed using data from 60 to 12 Å (final correlation coefficient of 0.35). Maps were calculated and averaged using data to 12 Å as a check on the procedure. The resolution was then extended to 6 Å and the process repeated (final correlation coefficient of 0.31). Cyclic averaging and solvent flattening of $2F_o - F_c$ maps (F_c being derived either from model phases or from inversion of a modified map) was used to refine the model phases in real space (program GAP; D.I.S. and J.M.G., unpublished data, CCP4 programs, and unpublished routines of R. Bryan) (final correlation coefficient 0.89).

Model building. A Cα model of the secondary structural units of the two molecules of VP3(T2) was built into the final 6 Å map. These units were refined as rigid bodies against data from 60 to 6 Å Bragg spacings, then 60–5 Å, and then to 3.8 Å (program X-PLOR, strict 30-fold n.c.s.). The n.c.s. operators were then re-refined, and real-space averaging resumed, using data to 3.8 Å. The $2F_o - F_c$ map was of sufficient quality for the unambiguous chain tracing and sequence alignment of both molecules of VP3(T2) (reciprocal space R -factors and correlation coefficients were 16.9% and 0.89, respectively, between F_o and F_c from the inverted-averaged/solvent-flattened map; see Table 1 for definition). The two molecules of VP3(T2) were built using FRODO³², and minor adjustments were made to the 13 monomers of VP7(T13).

Refinement. Meanwhile the data set had been extended to include some data to 3.5 Å (3,299,866 unique reflections, 21,510,811 total, $R_{\text{merge}} = 22.9\%$; Table 1). Data were sharpened by applying a B-factor of -44.5 Å², allowing a more sensible weighting of the data at different resolutions. The particle position and orientation were re-refined by rigid-body refinement in X-PLOR. The reciprocal space correlation coefficient after further phase refinement was 0.90, and the R -factor was 16.0%. The VP3(T2)–VP7(T13) model was then rebuilt into the final map and refined using X-PLOR (starting R -factor 34.4%). Positional and B-factor refinement led to an R -factor of 26.6% for all data between 100 and 3.5 Å (r.m.s. deviation in bond lengths and angles was 0.007 Å and 1.406°, respectively, r.m.s. deviation in B for bonded atoms was 5 Å²). Refinement improved the model as judged by the Ramachandran plot, with 84% of non-glycine residues in the refined model lying in the most favoured regions and 1% in disallowed regions³³. The figures of atomic structures were generated using BOBSCRIPT³⁴, and, as appropriate, rendered with RASTER3D^{35,36}.

Received 27 March; accepted 22 July 1998.

- Calisher, C. H. & Mertens, P. P. C. Taxonomy of African horse sickness virus. *Arch. Virol.* **13** (suppl.) (in the press).
- Holmes, I. H. et al. in *Virus Taxonomy: Classification and Nomenclature of Viruses* (eds Murphy, F. A. et al.) 208–237 (Springer, Wien and New York, 1998).
- Mertens, P. P. C. et al. Enhanced infectivity of modified bluetongue virus particles for two insect cell lines and for two *Culicoides* vector species. *Virology* **217**, 582–593 (1996).
- Furuichi, Y., Muthukrishnan, S., Tomasz, J. & Shatkin, A. J. Mechanism of formation of reovirus mRNA 5'-terminal blocked and methylated sequence, m7GpppGmpC. *J. Biol. Chem.* **251**, 5043–5053 (1976).
- Mertens, P. P. C., Burroughs, J. N. & Anderson, J. Purification and properties of virus particles, infectious subviral particles, and cores of bluetongue virus serotypes 1 and 4. *Virology* **157**, 375–386 (1987).
- Hewat, E. A., Booth, T. F., Loudon, P. T. & Roy, P. Three-dimensional reconstruction of baculovirus expressed bluetongue virus core-like particles by cryo-electron microscopy. *Virology* **189**, 10–20 (1992).
- Juuti, J. T., Ravantti, J. J. & Bamford, D. H. in *6th International Symposium on dsRNA Viruses* (eds Arias, C. F. & Lopez, S.) A1–A10 (Cocoyoc, Mexico, 1997).
- Prasad, B. V. V., Yamaguchi, S. & Roy, P. Three-dimensional structure of single-shelled bluetongue virus. *J. Virol.* **66**, 2135–2142 (1992).
- Grimes, J. M. et al. An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. *Structure* **5**, 885–893 (1997).
- Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
- Huisman, H., van Dijk, A. A. & Els, H. J. Uncoating of parental bluetongue virus to core and subcore particles in infected L cells. *Virology* **157**, 180–188 (1987).
- Loudon, P. T. & Roy, P. Assembly of five bluetongue virus proteins expressed by recombinant baculoviruses: inclusion of the largest protein VP1 in the core and virus-like particles. *Virology* **180**, 798–802 (1991).
- Burroughs, J. N., Grimes, J. M., Mertens, P. P. C. & Stuart, D. I. Crystallization and preliminary X-ray analysis of the core particle of bluetongue virus. *Virology* **210**, 217–220 (1995).

14. Burroughs, J. N. *et al.* Purification and properties of virus particles, infectious subviral particles, cores and VP7 crystals of African horsesickness virus serotype 9. *J. Gen. Virol.* **75**, 1849–1857 (1994).
15. Moss, S. R. & Nuttall, P. A. Subcore- and core-like particles of Broadhaven virus (BRDV), a tick-borne orbivirus, synthesized from baculovirus expressed VP2 and VP7, the major core proteins of BRDV. *Virus Res.* **32**, 401–407 (1994).
16. Fu, H., Bland, A. P., Wade-Evans, A. M., Mertens, P. P. C. & Mellor, P. S. Ultrastructural comparison of bluetongue virus replication in mammalian and *Culicoides* cells. *J. Gen. Virol.* (in the press).
17. Liddington, R. C. *et al.* Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**, 278–284 (1991).
18. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol.* **269**, 665–675 (1997).
19. Grimes, J. M., Basak, A. K., Roy, P. & Stuart, D. I. The crystal structure of bluetongue virus VP7. *Nature* **373**, 167–170 (1995).
20. Holm, L. & Sander, C. The FSSP database: fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.* **24**, 206–209 (1996).
21. Caston, J. R. *et al.* Structure of L-A virus: a specialized compartment for the transcription and replication of double-stranded rna. *J. Cell Biol.* **138**, 975–985 (1997).
22. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
23. Brookes, S. M., Hyatt, A. D. & Eaton, B. T. Characterizes of virus inclusion bodies in bluetongue virus-infected cells. *J. Gen. Virol.* **74**, 525–530 (1993).
24. Harrison, S. C. Protein interfaces and intersubunit bonding. The case of tomato bushy stunt virus. *Biophys. J.* **32**, 139–153 (1980).
25. Zhu, Y. *et al.* Details of the arrangement of the outer capsid of Rice Dwarf Phytoreovirus, as visualized by two-dimensional crystallography. *J. Virol.* **71**, 8899–8901 (1997).
26. Le Blois, H. & Roy, P. A single point mutation in the VP7 major core protein of bluetongue virus prevents the formation of core-like particles. *J. Virol.* **67**, 353–359 (1993).
27. Basak, A. K., Grimes, J. M., Gouet, P., Roy, P. & Stuart, D. I. Structures of orbivirus VP7: implications for the role of this protein in the viral life cycle. *Structure* **5**, 871–883 (1997).
28. Filman, D. J. *et al.* Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. *EMBO J.* **8**, 1567–1579 (1989).
29. Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H. & Fuller, S. D. Intermediates in the assembly pathway of the double-stranded RNA virus phi6. *EMBO J.* **16**, 4477–4487 (1997).
30. Otwinowski, Z. O. & Minor, W. in *Macromolecular Crystallography* (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic, San Diego, CA, 1997).
31. Brunger, A. T. *X-PLOR Version 3.1* (Yale Univ. Press, Connecticut, 1992).
32. Jones, T. A. in *Diffraction Methods for Biological Macromolecules* (eds Wyckoff, H. W., Hirs, C. H. W. & Tinsasheft, S. N.) 157–171 (Academic, Orlando, Florida, 1985).
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
34. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).
35. Bacon, D. & Anderson, W. F. A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220 (1988).
36. Merritt, E. A. & Murphy, M. E. P. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
37. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
38. Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* **204**, 1019–1029 (1988).
39. Frishman, D. & Argos, P. Knowledge-based secondary structure assignment. *Proteins* **23**, 566–579 (1995).
40. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
41. Garcia, K. C. *et al.* Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* **279**, 1166–1172 (1998).

Acknowledgements. We thank the SRS, ESRF and EMBL for their assistance in data collection, in particular the staff of ID2 at the ESRF; the disease security officers D. Goodridge, S. Williams, A. Meyer and P. Wilkinson; M. Pickford for technical assistance; R. Bryan, K. Measures and R. Esnouf for computing; S. Lee for assistance with figures; and P. Roy and E. Y. Jones for discussion. This work has been given long-term support by the BBSRC and MRC; R.M. is supported by a C. J. Martin fellowship from NHMRC Australia; P.P.C.M. and J.N.B. are supported by MAFF.

Correspondence and requests for materials should be addressed to D.I.S. (e-mail: dave@biop.ox.ac.uk). The coordinates have been deposited with the Protein Data Bank (accession number 2btv).

Dependence of magnetoresistivity on charge-carrier density in metallic ferromagnets and doped magnetic semiconductors

Pinaki Majumdar* & Peter B. Littlewood*†

* Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

† Cavendish Laboratory, Cambridge University, Madingley Road, Cambridge CB3 0HE, UK

Magnetoresistance—the field-dependent change in the electrical resistance of a ferromagnetic material—finds applications in technologies such as magnetic recording. Near and above the Curie point, T_c , corresponding to the onset of magnetic order, scattering of charge carriers by magnetic fluctuations can substantially increase the electrical resistance^{1,2}. These fluctuations can be suppressed³ by a magnetic field, leading to a negative magnetoresistance. Magnetic scattering might also have a role in the ‘colossal’ magnetoresistance observed in some perovskite manganese oxides^{4–6}, but is it not yet clear how to reconcile this behaviour with that of the conventional ferromagnetic materials. Here we show that, in generic models of magnetic scattering, the bulk low-field magnetoresistance (near and above T_c) is determined by a single parameter: the charge-carrier density. In agreement with experiment^{3,7,8}, the low-field magnetoresistance scales with the square of the ratio of the field-induced magnetization to the saturation magnetization. The scaling factor is $C \approx x^{-2/3}$, where x is the number of charge carriers per magnetic unit cell. Data from very different ferromagnetic metals and doped semiconductors are in broad quantitative agreement with this relationship, with the notable exception of the perovskite manganese oxides (in which dynamic lattice distortions complicate and enhance^{4,9–12} the effects of pure magnetic scattering). Our results might facilitate searches for new materials with large bulk magnetoresistive properties.

‘Colossal’ magnetoresistance (CMR) in some manganese oxides (such as the perovskite $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ and the pyrochlore $\text{Ti}_2\text{Mn}_2\text{O}_7$) has been found to accompany a transition from a metallic ferromagnetic low-temperature phase to a paramagnetic high-temperature phase⁴. The implication is that magnetic scattering of itinerant carriers from enhanced fluctuations near T_c has an important role. This prompts one to ask what are the principal differences from the behaviour in more familiar metallic ferromagnets such as Ni and Fe, other perovskite ferromagnets (for example, $\text{La}_{1-x}\text{Ca}_x\text{CoO}_3$), or doped magnetic semiconductors (for example, $\text{Ge}_{1-x}\text{Mn}_x\text{Te}$ and $\text{Eu}_{1-x}\text{Gd}_x\text{Te}$). The size of the magnetoresistance differs by at least an order of magnitude among these compounds, and each material might seem to demand a separate analysis, because the magnetic mechanisms can be very different. In the perovskite manganites and cobaltites as well as traditional ferromagnets such as Fe and Ni, the magnetic exchange itself arises from hopping of the conduction electrons (either the local ‘double exchange’ or the longer-range RKKY coupling), whereas in the Mn pyrochlores and magnetic semiconductors such as EuSe and CdCr_2Se_4 the magnetic coupling is mediated at least partly by bound electrons (superexchange). We argue that, provided that the transport remains metallic, the detailed model for the magnetism matters much less than a simple trend with carrier density.

Scattering from magnetic fluctuations might be a major component of the electrical resistance. Because the fluctuations at a

ferromagnetic transition lie near $q \sim 0$, they have usually only a weak effect on transport because it is primarily modes that can scatter backwards across the Fermi surface that are important—that is, with momenta of order $2k_F$ ($k_F = (3\pi^2 n)^{1/3}$ is the Fermi momentum in an electron gas of density n). The obvious and interesting exception is a low-density electron system $k_F a \ll 1$ (where a is the lattice constant), and it is in this low-density regime that the effects that we discuss are largest.

Our theoretical picture is a modest amplification of earlier work by Fisher and Langer¹; that work itself criticized and extended the approach of de Gennes and Friedel². Our results can be derived analytically in the Born scattering and Ornstein–Zernicke approximations that are not valid close to T_c , but the overall density-dependence survives in better theories with some modification to the temperature-dependence.

Whatever the precise origin of the magnetism, the spectrum of magnetic fluctuations can usually be captured by simple scaling forms even quite far from T_c . Our fundamental assumption is that these fluctuations generate nothing more than an effective static potential that scatters the carriers that give rise to transport.

The scattering rate, normalized to its high-temperature value (corresponding to uncorrelated spins), is given in the Born approximation by:

$$\tau^{-1}/\tau_0^{-1} \propto \int_0^\pi \sigma(\theta)(1 - \cos\theta)\sin\theta d\theta \quad (1)$$

where $\sigma(\theta)$ is the differential scattering cross section per magnetic spin, and θ is the scattering angle. The cross section, in turn, is given by:

$$\sigma(\theta) = \chi(q = 2k_F \sin(\theta/2)) \quad (2)$$

where q is the momentum transfer for this on-shell scattering, and $\chi(q)$ is the spin–spin correlation function.

The integral is dominated by large scattering angles, and to study trends it is sufficient to replace $\sigma(\theta)$ by $\sigma(\pi)$ in equation (1). Then the magnetoresistance is:

$$\Delta\rho/\rho \approx 1 - \chi(2k_F, T, H)/\chi(2k_F, T, 0) \quad (3)$$

at temperature T and magnetic field H .

We make a simple Ginzburg–Landau approximation for the low-field magnetic susceptibility in the paramagnetic state, namely $\chi^{-1}(q, T, H) \propto A(T) + (q\xi_0)^2 + (m(H)/m_{\text{sat}})^2$, where $A(T)$ vanishes at T_c . The correlation length $\xi(T) = \xi_0 A(T)^{1/2}$ diverges as $T \rightarrow T_c$; the corresponding divergence in χ is cut off either by the momentum q or by the field-induced magnetization $m(H)$. The scale m_{sat} is comparable to the saturation magnetization in a large field. Using equation (3), we get, at low fields:

$$\Delta\rho/\rho = C(m/m_{\text{sat}})^2 \approx (1/2k_F\xi_0)^2(m/m_{\text{sat}})^2 \quad (4)$$

provided that $k_F\xi(T) \gg 1$. This formula depends solely on elementary parameters of the magnetic system: the bare correlation length (of order the separation between magnetic moments) and the saturation magnetization. We shall discuss the validity of the approximations made to derive equation (4) below, but first compare with available experiments, listed in Table 1 and plotted in Fig. 1.

The data shown are a reanalysis of previously published data, which are limited because we require measurements of both the nonlinear field-dependent magnetization and the magnetoresistance, as well as an unambiguous determination of the carrier density. We have used data in the range 1.1 – $1.5 T_c$. A few authors^{3,7,8} have recognized the scaling of magnetoresistance with magnetization, and report values of the coefficient C , which we have used. Because our theory should apply only to metallic conduction, we have (with the exception of the perovskite manganites) not included data for cases in which metallic conduction was not observed.

To determine the overall magnetic component of the resistivity, we have used either the high-field (saturated) magnetoresistivity,

where available, or have estimated a smooth background resistivity curve through T_c . For cases in which the magnetic scattering is a small part of the total (where C is also small) this procedure might introduce a substantial error. But in all cases where we estimate $C > 2$, the magnetic contribution to the resistivity was inferred to be more than 50% of the total resistance. The carrier concentration was derived from the nominal chemical composition, except for $\text{Ti}_2\text{Mn}_2\text{O}_7$ (ref. 8), $\text{Ge}_{0.99}\text{Mn}_{0.01}\text{Te}$ (ref. 13) and $\text{Eu}_{1-x}\text{Gd}_x\text{Se}$ (ref. 3), for which Hall measurements were used. ξ_0 was taken to be the separation between magnetic ions, that is approximately the lattice constant in all cases but $\text{Ge}_{0.99}\text{Mn}_{0.01}\text{Te}$.

The advertised trend—increasing C with decreasing carrier concentration—is clear, with the exception of the perovskite manganese oxides, discussed below. Note that the isostructural $\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$, for which the usual double-exchange model is believed to be adequate, has appropriately small value of C at the concentrations ($x = 0.2$ and 0.3) reported¹⁴. The pyrochlore manganite $\text{Ti}_2\text{Mn}_2\text{O}_7$ fits into the main group of materials and not with the perovskites, consistent with its Mn^{4+} valence and correspondingly weak electron-lattice coupling. Another interesting comparison is between the doped semiconductors: $\text{Ge}_{0.99}\text{Mn}_{0.01}\text{Te}$ has approximately the same carrier density per unit cell as the spinel $\text{Ga}_{0.02}\text{Cd}_{0.98}\text{Cr}_2\text{Se}_4$, but differs by a factor of 20 in the scaled magnetoresistance C (a difference explained by the dilution of the magnetic ions in the GeMnTe so that the values of $n\xi_0^3$ differ by two orders of magnitude between the two compounds).

In the perovskite manganites, it has been known for some time that the carrier density is not the sole determinant of the magnetoresistance⁹. Our analysis of the data provides further evidence for the already preponderant conclusion that other physical processes are at work in the perovskite manganites, now established to be dynamical electron-lattice interactions that are strong in a range of compositions near $x = 0.1-0.2$ and drive the paramagnetic phase to be non-metallic^{4,10-12}. (It should be evident that the decrease in resistance at T_c by two orders of magnitude in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ between $x = 0.15$ and $x = 0.3$ cannot be captured by a model with spin scattering alone.) Notice that for the metallic Sr-doped manganites ($x = 0.3$ or 0.4), in which the lattice coupling becomes weaker, the magnetoresistance is again small and the temperature dependence of the resistivity is characteristic of a metal.

Several approximations are made in our elementary calculation that deserve note. The use of the Ginzburg–Landau approximation for the correlation function will lead to a cusp in the resistance at T_c which is unphysical. A better theory¹ for χ leads to a peak in τ^{-1}

Table 1 Sources of data shown in Fig. 1

	x or $n\xi_0^3$	C	Reference
$\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$	0.25	6–8	22
	0.15	4	
	0.175	4	
$\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$	0.20	4	7
	0.30	2	
	0.40	1	
$(\text{La}_{1-y}\text{Y}_y)_{2/3}\text{Ca}_{1/3}\text{MnO}_3$	0.33	0.8	
	0.33	1.5	23
	0.33	2	
$\text{Ti}_2\text{Mn}_2\text{O}_7$	0.001–0.005	15	8
$\text{Ga}_{0.02}\text{Cd}_{0.98}\text{Cr}_2\text{Se}_4$	0.02	7–8	24
$\text{Eu}_{0.95}\text{Gd}_{0.05}\text{Se}$	0.003	5–15	3
$\text{Eu}_{0.99}\text{Gd}_{0.01}\text{Se}$	0.0002	14	(data below T_c)
$\text{Ge}_{0.99}\text{Mn}_{0.01}\text{Te}$	4.5	0.4	13
$\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$	0.2	1.2	14
	0.3	0.5	
Gd	1	1.2 ± 0.3	A. P. Ramirez (unpublished observations)
FeCr_2S_4	0.02–0.05	2.5–3	25
$\text{Fe}_{0.5}\text{Cu}_{0.5}\text{Cr}_2\text{S}_4$	0.03–0.07	2.5–3	25

above T_c (albeit very close if $2k_F \ll \xi_0$), but with a similar magnitude and the same density dependence as our analysis above. For this reason, we restricted our analysis to data taken in the range $1.1T_c < T < 1.5T_c$. We remark that one obtains $dp/dT < 0$ if k_F is small enough, despite the metallic conduction.

Although the magnetoresistance (equation (4)) is (in our approximation) independent of the coupling constant between the conduction electrons and the magnetic fluctuations, the resistance itself is not. If the coupling constant were large enough, the Born calculation would predict a mean free path l that could become arbitrarily short; the result is meaningless if it predicts $k_F l < 1$, as it must at low enough density. However, numerical calculations¹⁵ have shown unambiguously that even for large exchange coupling of the spins to the carriers, spin scattering by itself can lead to localization only far into the tail of the band, for carrier densities less than a critical value n_c found to be smaller than 0.005. The perturbation theory that we have presented might be qualitatively accurate throughout a wide density regime. Only at very low density will the picture of scattering be replaced by one of true localization.

The above-mentioned results^{10,15} are in contradiction with a dynamical field theory in infinite dimension¹⁶ that claimed that strong coupling by itself could lead to incoherent transport, generating both large resistance and magnetoresistance. In that work, the weak coupling value of the coefficient C was stated to be unity, which is not correct at least in finite dimension; and the large value of the electronic scattering rate (comparable to the electronic bandwidth) found was considerably greater than the value found by calculation in three dimensions^{10,15}, and other dynamical mean field theories using a more physical density of states¹⁷.

Another localizing effect at low density will be a tendency to form magnetic polarons, in which carrier hopping polarizes a local ferromagnetic cluster of moments and thereby self-traps. Polarons can exist as well-defined non-overlapping entities provided that $n\xi^3 \ll 1$ (or equivalently $k_F \xi \ll 1$). This regime can be reached only at temperatures somewhat above T_c , even if k_F is small; at temperatures higher still, thermal fluctuations will destroy the polaron (ref. 18; and P.M. and P.B.L., unpublished observations). Either kind of

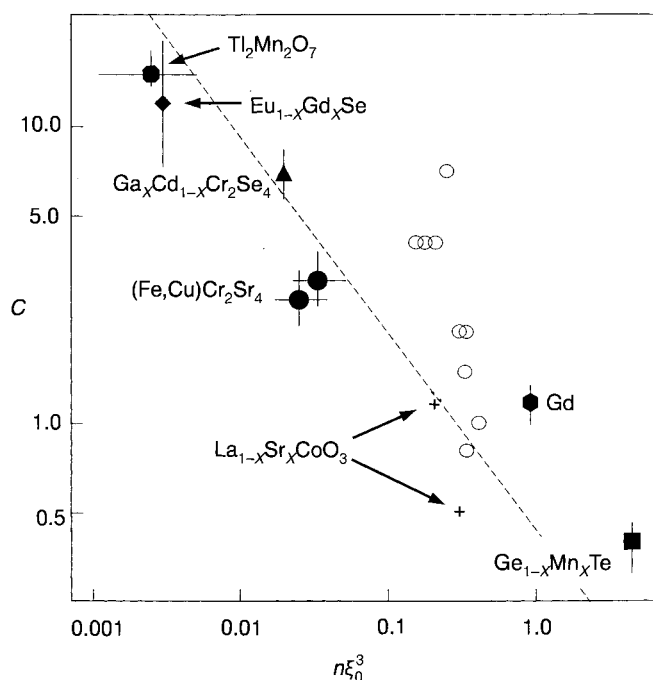


Figure 1 Low-field magnetoresistance $C = (\Delta\rho/\rho)/(m(H)/m_{\text{sat}})^2$ as a function of the carrier density scaled to the magnetic lattice spacing. Open circles, perovskite manganites; solid symbols, from other metallic ferromagnets. The dashed line shows the $n^{-2/3}$ trend predicted by Born scattering.

localizing effect will lead to saturation of the magnetoresistance. Yet all the data except for the perovskite manganites and the strongly localized $\text{Eu}_{0.99}\text{Gd}_{0.01}\text{Se}$ ($n = 0.0002$ per unit cell) are on samples that satisfy $k_F l > 1$, indicating the weakly localizing nature of pure spin scattering, as expected.

At low carrier densities, one should be wary of magnetization-dependent bandshift, because the exchange energy might be comparable to the Fermi energy. When the carriers are introduced by doping, this is not important, but for semimetals ($\text{Ti}_2\text{Mn}_2\text{O}_7$ may be an example^{19,20}) and mixed valent compounds²¹ the carrier density itself might become field- and temperature-dependent.

We conclude that on elementary theoretical grounds and with some support from the limited data available, the very simplest Born approximation gives a good account of the trends in bulk magnetoresistance over two decades in density, in materials and in models of magnetism that have otherwise very little in common. The most evident deviation from a simple scaling law is in the perovskite manganites, for which there is overwhelming evidence that their 'colossal' magnetoresistance arises from the localization of carriers by lattice distortions that are themselves coupled to the magnetic degrees of freedom.

This organizing principle suggests that insulating ferromagnets, lightly doped into the metallic regime, are a promising class of materials in which to search for large bulk magnetoresistance. As a final cautionary note, we also point out that the scaling shown in equation (4) and in Fig. 1 nevertheless implies rather small values of the absolute bulk magnetoresistance at the millitesla fields of technological interest. □

Received 27 April; accepted 9 July 1998.

1. Fisher, M. E. & Langer, J. S. Resistive anomalies at magnetic critical points. *Phys. Rev. Lett.* **20**, 665–668 (1968).
2. de Gennes, P. G. & Friedel, J. Anomalies de résistivité dans certains métaux magnétiques. *J. Phys. Chem. Solids* **4**, 71–77 (1958).
3. von Molnar, S. & Methfessel, S. Giant negative magnetoresistance in ferromagnetic $\text{Eu}_{1-x}\text{Gd}_x\text{Se}$. *J. Appl. Phys.* **38**, 959–964 (1967).
4. Ramirez, A. P. Colossal magnetoresistance. *J. Phys. Condens. Matt.* **9**, 8171–8199 (1997).
5. Mathur, N. Not just a load of bolometers. *Nature* **390**, 229–231 (1997).
6. Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 414–417 (1994).
7. Urushibara, A. *et al.* Insulator–metal transition and giant magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 14103–14109 (1995).
8. Shimikawa, Y., Kubo, Y. & Manako, T. Giant magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$ with the pyrochlore structure. *Nature* **379**, 53–55 (1996).
9. Hwang, H. Y., Cheong, S.-W., Radaelli, P. G., Marezio, M. & Batlogg, B. Lattice effects on the magnetoresistance in doped LaMnO_3 . *Phys. Rev. Lett.* **75**, 914–917 (1995).
10. Millis, A. J., Littlewood, P. B. & Shraiman, B. I. Double exchange alone does not explain the resistivity of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **74**, 5144–5147 (1995).
11. Millis, A. J., Shraiman, B. I. & Mueller, R. Dynamic Jahn–Teller effect and colossal magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. Lett.* **77**, 175–178 (1996).
12. Röder, H., Zang, J. & Bishop, A. R. Lattice effects in colossal magnetoresistance perovskites. *Phys. Rev. Lett.* **76**, 1356–1359 (1996).
13. Cochrane, R. W., Hedgcock, F. T. & Ström-Olsen, J. O. Exchange scattering in a ferromagnetic semiconductor. *Phys. Rev. B* **8**, 4262–4266 (1973).
14. Yamaguchi, S., Taniguchi, H., Takagi, H., Arima, T. & Tokura, Y. Magnetoresistance in metallic crystals of $\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$. *J. Phys. Soc. Jpn* **64**, 1885–1888 (1995).
15. Li, Q., Zang, J., Bishop, A. R. & Soukoulis, C. M. Charge localization in disordered colossal magnetoresistance manganites. *Phys. Rev. B* **56**, 4541–4544 (1997).
16. Furukawa, N. Magnetoresistance of the double-exchange model in infinite dimension. *J. Phys. Soc. Jpn* **64**, 2734–2737 (1995).
17. Green, A. C. M., Edwards, D. M. & Kubo, K. A generalized coherent potential approximation for double exchange systems. Preprint (1998).
18. Kasuya, T., Yanase, A. & Takeda, T. Stability condition for the magnetic polaron in a magnetic semiconductor. *Solid State Commun.* **8**, 1543–1546 (1970).
19. Singh, D. J. Magneto-electronic effects in pyrochlore $\text{Ti}_2\text{Mn}_2\text{O}_7$: role of $\text{Ti}^3\text{–O}$ covalency. *Phys. Rev. B* **55**, 313–316 (1997).
20. Ventura, C. I. & Alascio, B. Intermediate valence model for the colossal magnetoresistance in $\text{Ti}_2\text{Mn}_2\text{O}_7$. *Phys. Rev. B* **56**, 14533–14540 (1997).
21. Batlogg, B. Valence changes in TiMnSe by alloying with TiTe and EuSe . *Phys. Rev. B* **23**, 650–663 (1981).
22. Schiffer, P., Ramirez, A. P., Bao, W. & Cheong, S.-W. Low-temperature magnetoresistance and the magnetic phase diagram of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **75**, 3336–3339 (1995).
23. Martinez, B. *et al.* Spin–disorder band scattering and localization in magnetoresistive $(\text{La}_{1-x}\text{Y}_x)_{0.5}\text{Ca}_{0.5}\text{MnO}_3$ perovskites. *Phys. Rev. B* **54**, 10001–10007 (1996).
24. Haas, C., van Run, A. M. J. G., Bongers, P. F. & Albers, W. The magnetoresistance of n-type CdCr_2Se_4 . *Solid State Commun.* **5**, 657–661 (1967).
25. Ramirez, A. P., Cava, R. J. & Krajewski, J. Colossal magnetoresistance in Cr-based chalcogenide spinels. *Nature* **386**, 156–159 (1997).

Acknowledgements. We thank G. Aeppli, B. Batlogg, H. Hwang, A. Millis and A. Ramirez for extensive discussions. Work done at Cambridge University was supported by the EPSRC.

Correspondence and requests for materials should be addressed to P.B.L. (e-mail: peter.littlewood@phy.cam.ac.uk).

Electric-field-enhanced crystallization of amorphous silicon

Jin Jang*, Jae Young Oh*, Sung Ki Kim*, Young Jin Choi*, Soo Young Yoon† & Chae Ok Kim†

* Department of Physics, Kyung Hee University, Dongdaemoon-ku, Seoul 130-701, Korea

† Department of Physics, Han Yang University, Sungdong-ku, Seoul 133-791, Korea

Thin films of polycrystalline silicon are of great importance for large-area electronic applications, providing, for example, the switching electronics in many flat-panel displays. Polycrystalline silicon is typically produced by annealing films of amorphous silicon¹ that have been deposited from the vapour phase, and much research is focused on lowering the crystallization temperature. It is known that the solid-phase crystallization temperature of amorphous silicon can be reduced by the addition of certain metals², such as nickel³. Here we show that the rate at which this metal-induced crystallization takes place is markedly enhanced in the presence of an electric field. For example, the crystallization time at 500 °C decreases from 25 hours to 10 minutes on application of a modest (80 V cm^{−1}) electric field. No residual amorphous phase can be detected in the films. A thin-film transistor fabricated from such a film exhibits a field-effect mobility of 58 cm² V^{−1} s^{−1}, thereby demonstrating the practical utility of these materials.

Hydrogenated amorphous silicon (a-Si:H) films of 400-nm thickness were deposited by plasma-enhanced chemical vapour deposition using a mixture of SiH_4 and H_2 at the substrate temperature of 275 °C. A 0.03-Å-thick Ni layer was deposited over the whole a-Si:H by sputtering, and then 30-nm-thick Mo electrodes (with a parallel gap separation of 2.5 cm) were deposited using a shadow mask. The samples were crystallized at 500 °C in the presence of an electric field applied between the two electrodes in vacuum.

The detailed fabrication process of the polycrystalline-silicon thin-film transistors (poly-Si TFTs) was as follows⁴. A very thin layer (0.03 Å) of Ni was deposited on the 400-nm-thick a-Si:H on glass; the a-Si:H was then crystallized at a temperature of 500 °C for 10 min in an electric field of 80 V cm^{−1}. The poly-Si layer was patterned, and then 250-nm-thick SiN_x and 30-nm-thick a-Si:H layers were deposited sequentially on the poly-Si. For SiN_x deposition, the gas flow rates (in cm³ STP min^{−1}) of He , SiH_4 and NH_3 were fixed at 50, 1 and 30, respectively; for a-Si:H deposition, the gas flow rate of SiH_4 was fixed at 1. The top two layers of a-Si:H and SiN_x were patterned in order to produce a channel, followed by ion doping on the top a-Si:H and the side poly-Si with 1% PH_3 diluted in H_2 plasma, using an ion doping system. The ion acceleration voltage, gas pressure and doping temperature were fixed at 6 kV, 5×10^{-3} torr and 230 °C, respectively. Under these conditions the ion dose was fixed at 5×10^{16} cm^{−2} to produce the n⁺ layer. Then 20-nm-thick Ni was deposited by sputtering, and annealed at 250 °C for 1 h to form Ni silicide in the source/drain/gate contacts. The unreacted nickel on the silicide was etched away in a ($\text{HNO}_3 + \text{HCl}$) solution⁵. The ratio of width to channel length of the poly-Si TFT was 80 μm/17 μm.

Figure 1 shows transmission electron microscopy (TEM) bright-field images of the partially crystallized poly-Si region (Fig. 1a) and the completely crystallized poly-Si region (Fig. 1b), and their electron-diffraction patterns. The grain size of the poly-Si produced by metal-induced crystallization (MIC) is seen to be >100 nm from

the TEM, indicating a good crystalline quality. In the diffraction pattern of the partially crystallized region, a grey disk is observed together with the spots for Si {011}. This disk is due to remaining a-Si, and the {011} planes are the results of laterally grown crystallites to the {111} directions.

The progress of silicide-mediated crystallization of a-Si films (by annealing at 500 °C in an electric field of 80 V cm⁻¹) is shown by the sequence of TEM bright-field images in Fig. 2. Precipitates of NiSi₂ are easily formed at 500 °C, as shown in Fig. 2a; this is expected, because precipitates of this compound are known to form at temperatures as low as 350 °C on a-Si (ref. 6). The precipitates are regular octahedral bounded by eight {111} faces⁷. The precipitates of NiSi₂ aggregate together, and become bigger precipitates, as shown in Fig. 2b. In Fig. 2c, a few needles in the early stage of growth from the bigger precipitates can be seen. NiSi₂ precipitates are clearly visible within the a-Si film as well as in the crystalline region. As shown in the figure, crystallization proceeds in the directions of {111} after reaching a critical size (~80 nm in the present work). As

shown in Fig. 1a, the electron-diffraction pattern indicates that the orientations of the needle-like crystallites are in the {011} directions. In Fig. 2c, the angle between the two needle-like crystallites, branching from one needle, is ~70.5° in one crystalline region because electron diffraction reveals that the single-crystal Si exhibits a {011} orientation with respect to the film surface normal. TEM studies revealed that the nucleation of Si occurs on the {111} face of an individual precipitate. The orientation of the NiSi₂ precipitate within the a-Si thin film determines the orientation of the growth direction of Si needles: all needles grow in {111} directions. The crystallites grow in the directions of {111} on an a-Si surface with {011} orientation with respect to the film surface normal; in this case only 70.5° is available in the diamond-structure⁸.

Extensive crystallization of two separately grown crystalline regions is seen in Fig. 2d; in this case the growth direction in one crystalline region is different from that in the second region because the two regions were independently grown from different nucleation sites. In Fig. 2e, many independently grown crystalline regions appear in the amorphous network. The needle-like crystallite grows as a result of the diffusion of Ni atoms in NiSi₂ towards the amorphous silicon interface because the free energy of an Ni atom at the interface between NiSi₂ and crystalline Si is higher than that of an Ni atom between NiSi₂ and a-Si (ref. 7).

With increasing crystallization time, the crystalline regions meet, and thus the amorphous silicon is fully crystallized. The interfaces between the different crystalline regions form clear grain boundary

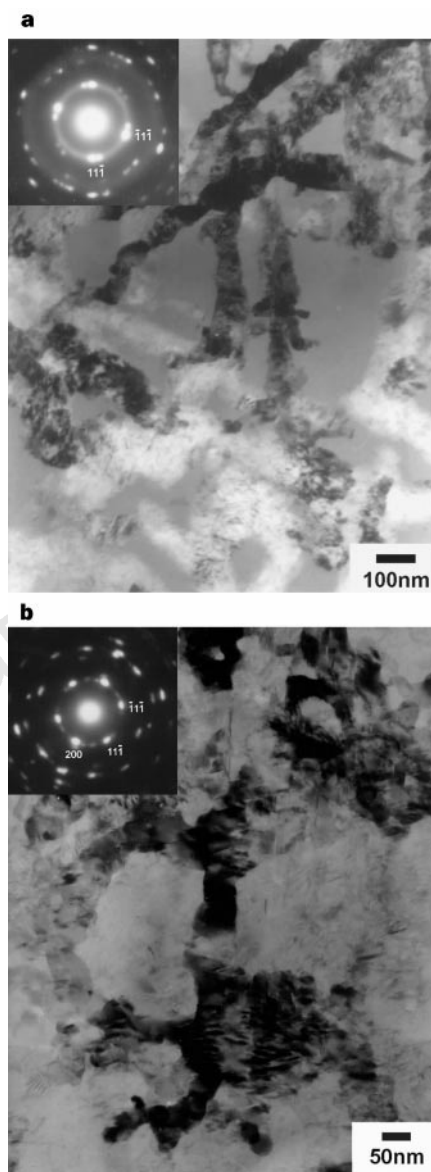


Figure 1 TEM bright-field images of partially crystallized (a) and completely crystallized (b) regions of Si films. These films were crystallized at 500 °C in an electric field of 80 V cm⁻¹; a shows a film after 5 min, b shows a film after 10 min. Analyses of the TEM images and electron-diffraction patterns of the completely crystallized region showed that it contained no amorphous phase.

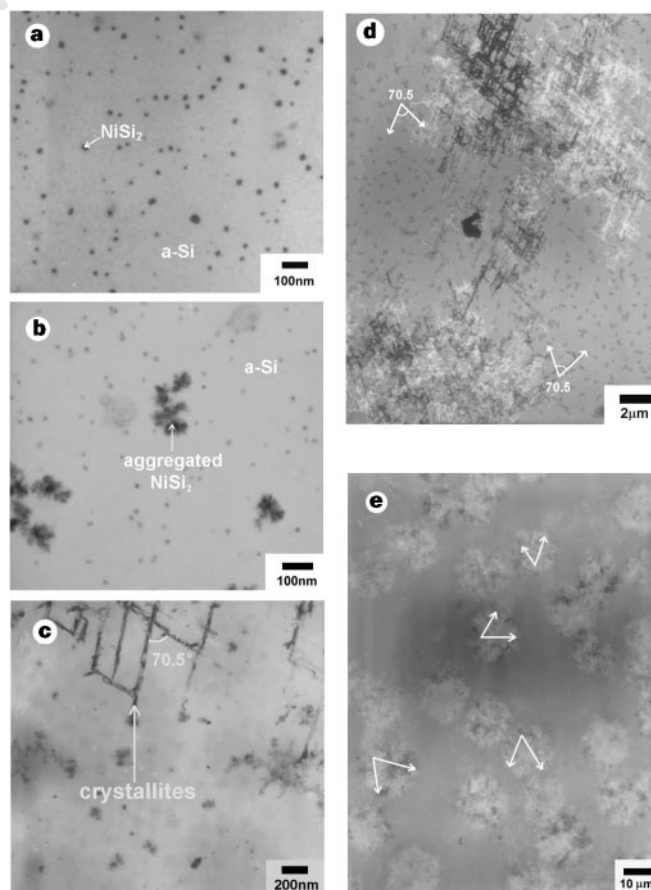


Figure 2 TEM bright-field images showing the progress of the silicide-mediated crystallization of a-Si films. These films were annealed at 500 °C in an electric field of 80 V cm⁻¹: a, formation of silicide precipitates after 20 s annealing; b, aggregation of NiSi₂ silicide precipitates after 40 s annealing; c, growth of needle-like Si crystallites after 60 s annealing; d, extensive growth of needle-like crystallites after 120 s annealing; e, the many crystalline regions in the amorphous network also after 120 s annealing.

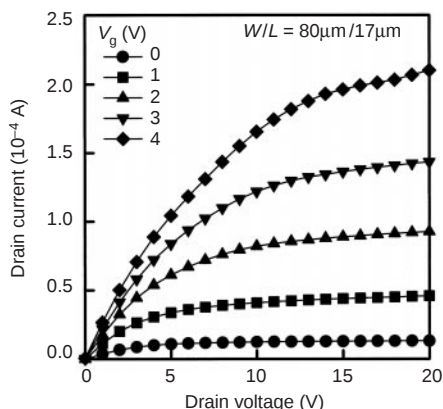


Figure 3 Drain current-drain voltage characteristics of the poly-Si TFT. The poly-Si layer was crystallized at 500 °C for 10 min in an electric field of 80 V cm⁻¹.

regions having defect states that are able to trap carriers and thus the mobility of carriers will decrease. But the boundary between two needles formed at 70.5° in one crystalline region does not create defect states because of the continuous formation of lattice from the mother needle. Therefore, there can be no grain-boundary states within one crystalline region. This is the advantage of our crystallization method. Although all precipitates could act as nucleation sites, substantial growth of Si can occur for {011} orientated precipitates. The {100} and {111} orientated precipitates exhibit {111} planes that intersect the upper or lower surface of the a-Si film. Therefore, the crystallization proceeds in {111} directions laterally from precipitates with {011} orientation with respect to the film surface normal⁷.

We note that earlier observations⁶ of Ni-induced crystallization of a-Si (without an electric field applied) revealed that the onset temperature for the crystallization of a-Si is significantly reduced in the presence of NiSi₂ precipitates and occurs at 500 °C. The crystallization is mediated by the migration of NiSi₂ precipitates. In the initial state the nucleation occurs at the {111} faces of the individual precipitates. The orientation of NiSi₂ precipitates within the a-Si determines both the orientation of the initial crystalline structure and the subsequent growth direction of the crystallites. All kinds of crystallites appear to grow laterally in the {111} directions⁷.

The Ni atoms in NiSi₂ can be negatively charged because of the negative Mulliken charge—that is, the net charge of Ni in a silicon network (Y. H. Lee, personal communication). The diffusion coefficient of Ni atoms therefore increases with the electric field. Enhancement of the diffusion coefficient of Ni atoms in the presence of an electric field has been found⁸ in the formation of Ni silicide: the formation of this compound is accelerated in the presence of an electric field in the Ni/a-Si:H system⁸.

The drain current-drain voltage ($I_d - V_d$) characteristics of the poly-Si TFT is shown in Fig. 3. The absence of the current crowding effect at low drain voltages confirms the good ohmic contacts obtained using the Ni silicide layer. The sheet resistance of the Ni silicide layer was found to be 3 Ω per square.

Figure 4 shows the channel conductance for the MIC poly-Si TFT; this device shows a field effect mobility of 58 cm² V⁻¹ s⁻¹ and a threshold voltage of -0.8 V. The field effect mobility (μ_{fe}) and threshold voltage (V_{th}) were obtained from the channel conductance (g_0) as follows⁹:

$$g_0 = \frac{I_d}{V_d} \Big|_{V_d \rightarrow 0} = C_i \mu_{fe} \frac{W}{L} (V_g - V_{th})$$

where C_i and W/L are respectively the capacitance of the gate insulator and the ratio of channel width to channel length of the TFT.

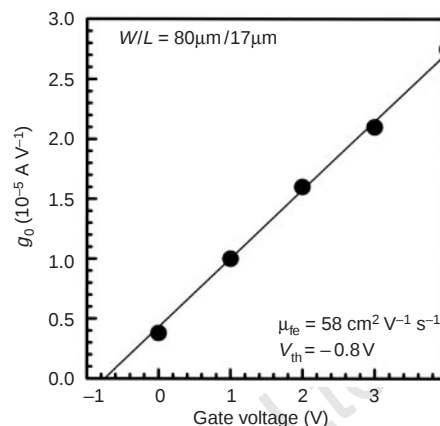


Figure 4 Channel conductance as a function of gate voltage for the MIC poly-Si TFT.

We consider that the electric-field-enhanced MIC technique could be widely applied to crystallize amorphous materials. □

Received 12 May; accepted 15 July 1998.

1. Masaki, Y., LeComber, P. G. & Fitzgerald, A. G. Solid phase crystallization of thin films of Si prepared by plasma enhanced chemical vapor deposition. *J. Appl. Phys.* **74**, 129–134 (1993).
2. Konno, T. J. & Sinclair, R. Metal-contact-induced crystallization of semiconductors. *Mater. Sci. Eng. A* **179/180**, 426–432 (1994).
3. Kawazu, Y., Kudo, H., Onari, S. & Arai, T. Low temperature crystallization of hydrogenated amorphous silicon induced by nickel silicide formation. *Jpn J. Appl. Phys.* **29**, 2698–2704 (1990).
4. Ryu, J. I., Kim, H. C., Kim, S. K. & Jang, J. A novel self-aligned polycrystalline silicon thin-film transistor using silicide layers. *IEEE Electron Device Lett.* **18**, 272–274 (1997).
5. Murarka, S. P. *Metallization: Theory and Practice for VLSI and ULSI* Ch. 2 (Butterworth-Heinemann, Boston, 1993).
6. Camareata, R. C., Thompson, C. V., Hayzelden, C. & Tu, K. N. Silicide precipitation and silicon crystallization in nickel implanted amorphous silicon thin films. *J. Mater. Res.* **5**, 2133–2138 (1990).
7. Hayzelden, C. & Batstone, J. L. Silicide formation and silicide-mediated crystallization of nickel-implanted amorphous silicon thin films. *J. Appl. Phys.* **73**, 8279–8289 (1993).
8. Murakami, H., Ono, K. & Takai, H. Effect of electric field on silicide formation. *Appl. Surf. Sci.* **117/118**, 289–293 (1997).
9. Sze, S. M. *Physics of Semiconductor Devices* 2nd edn, Ch. 8 (Wiley, New York, 1981).

Acknowledgements. This work was supported by the Korean G-7 project.

Correspondence and requests for materials should be addressed to J.J. (e-mail: jjang@nms.kyunghee.ac.kr).

Molecular-scale mechanisms of crystal growth in barite

Carlos M. Pina, Udo Becker, Peter Risthaus, Dirk Bosbach & Andrew Putnis

Institut für Mineralogie, Universität Münster, Corrensstrasse 24, 48149 Münster, Germany

Models of crystal growth have been defined by comparing macroscopic growth kinetics with theoretical predictions for various growth mechanisms^{1,2}. The classic Burton–Cabrera–Frank (BCF) theory³ predicts that spiral growth at screw dislocations will dominate near equilibrium. Although this has often been observed^{2,4}, such growth is sometimes inhibited^{4,5}, which has been assumed to be due to the presence of impurities⁶. At higher supersaturations, growth is commonly modelled by two-dimensional nucleation on the pre-existing surface according to the ‘birth and spread’ model⁷. In general, the morphology of a growing crystal is determined by the rate of growth of different crystallographic faces, and periodic-bond-chain (PBC) theory^{8,9} relates this morphology to the existence of chains of strongly bonded ions in the structure. Here we report tests of such models for the growth of barite crystals, using a combination of *in situ* observations of growth mechanisms at molecular resolution with

the atomic force microscope^{10,11} and computer simulations of the surface attachment of growth units. We observe strongly anisotropic growth of two-dimensional nuclei with morphologies controlled by the underlying crystal structure, as well as structure-induced self-inhibition of spiral growth. Our results reveal the limitations of both the BCF and PBC theories in providing a general description of crystal growth.

Observations of crystal growth mechanisms were made in a fluid cell of a DI Multi-Mode atomic force microscope (AFM), using freshly cleaved, optically clear natural barite crystals as substrates. The surfaces studied were (001) and (210), the morphologically most important in barite¹². We used X-ray diffraction to confirm the cleavage surfaces as (001) and (210) and to define crystallographic directions in the plane of the substrate. The unit cell of the barite structure¹³ ($a = 8.87 \text{ \AA}$, $b = 5.45 \text{ \AA}$, $c = 7.15 \text{ \AA}$, space group $Pnma$) consists of two BaSO_4 layers parallel to (001), each layer related by a 2_1 screw diad axis parallel to c .

The cleaved (001) surface typically shows unit-cell high cleavage steps, and in Fig. 1a the cleavage surface intersects a screw dislocation line. At this point, such a cleavage step terminates. At the start of the experiment, deionized water was passed over the crystal, causing slight dissolution and the separation of the cleavage steps into half-steps (Fig. 1a). The presence of the screw dislocation allows spiral growth to be followed from the earliest stage and compared with two-dimensional nucleation.

A few seconds after injecting slightly supersaturated BaSO_4 solution, crystal growth begins. The cleavage steps migrate normal to their length and spiral growth begins at the dislocation

core (Fig 1b, c). A new BaSO_4 layer spreads away from the core to one side but does not continue around the spiral as in classical descriptions of spiral growth⁵. Instead, the spiral becomes increasingly tightly wound around the core with very little lateral growth. Subsequently, a small number of two-dimensional nuclei, one BaSO_4 layer thick ($3.6 \text{ \AA}$), are formed on the surface. These nuclei are bounded by two straight edges parallel to $\langle 120 \rangle$ directions and a curved edge defining a sector shape (Fig. 1d). The nuclei on each alternate BaSO_4 layer are oriented in opposite directions, with those on a newly grown layer related by a 2_1 diad axis to those on the original cleavage surface (Fig. 1e). The growth of these nucleated islands is highly anisotropic, with growth normal to the curved edges being of the order of 10 times faster than growth normal to the straight edges. A fresh injection of more supersaturated solution into the fluid cell results in a higher nucleation density (Fig. 1f), but the lateral growth of the spiral remains very much slower than that of the two-dimensional nuclei, forming a small growth hillock.

The microtopographic features of the nuclei and the spiral on the barite (001) face indicate strong structural control of the growth process. Observations made on $\{210\}$ cleavage surfaces of the same barite crystals show a very different growth morphology (Fig. 2). On the (210) barite surface, the nuclei grow as long needles that are one molecular layer thick ($3.4 \text{ \AA}$) along the $\langle 120 \rangle$ directions. The growth rate is also highly anisotropic along the length of the needles, with growth being ~ 10 times faster along the $[120]$ direction than in the opposite $[\bar{1}20]$ direction. Figure 3a shows the relevant crystallographic orientations on both barite surfaces.

These observations raise important questions relating to the

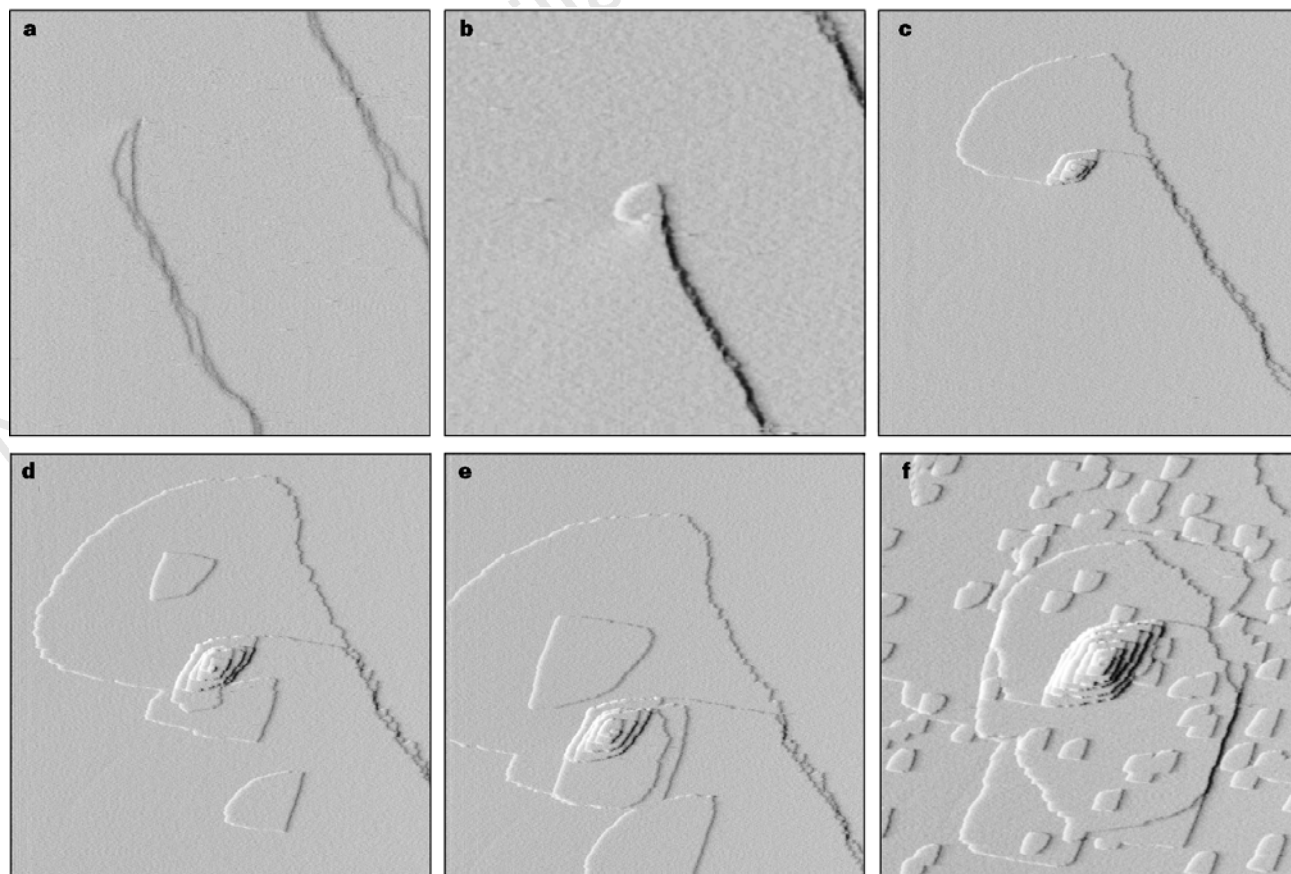


Figure 1 The crystal growth sequence on an (001) cleavage plane in a BaSO_4 solution with a supersaturation value of $a(\text{Ba}^{2+})/a(\text{SO}_4^{2-})/K_{sp} = 12$. (Here a is the ion activity and K_{sp} is the solubility product.) **a–e**, Growth sequence over 60 min. Spiral growth begins at the dislocation core with the growth of a molecular layer of BaSO_4 (thickness, $3.6 \text{ \AA}$), but subsequent lateral growth around the spiral is restricted. After ~ 40 min, during which fresh fluid is pumped over the sample,

two-dimensional nucleation begins (**d**), with a nucleation density of ~ 0.12 nuclei per μm^2 . Note the opposite orientation of the sector-shaped nuclei. **f**, After 85 min of growth, a fresh injection of solution with a supersaturation value of 26 results in a greatly increased nucleation rate (~ 22 nuclei per μm^2). The image area in each case is $2 \mu\text{m} \times 2 \mu\text{m}$.

crystal growth process, none of which can be adequately addressed using current theories. What determines the shape of the two-dimensional nuclei, why is their growth so anisotropic, and what inhibits the lateral growth of the spirals?

The prediction of crystal growth morphology is based on the concept that crystal faces contain different numbers of chains of nearest-neighbour bonds, or PBCs (periodic bond chains)^{8,14}. PBC analysis can also provide information about the surface morphologies on F-faces¹⁵, which contain two or more such PBCs and can grow by a layer mechanism. Both (001) and (210) are F-faces and PBC theory explains the layer-by-layer mechanism and the thickness of the growth layers: half-unit cell for (001) and one unit cell for (210) faces^{9,16}. In addition, needles on (210) and the straight edges of the sector-shaped nuclei on (001) are parallel to $\langle 120 \rangle$, which are directions of strong PBCs. However, as PBC theory was designed to deal with periodic systems and not with the attachment of single molecules at specific surface sites, it is necessary to apply molecular modelling techniques to explain the anisotropy of the growth along the PBCs and the existence of the curved edge.

We used ionic models incorporated in Cerius2 (Molecular Simulations, BIOSYM) and GULP (J. Gale, Imperial College, London) with potentials similar to those derived in ref. 12, but slightly modified to model the lattice energy of barite better. All values for attachment energies were corrected for the hydration energies of the respective ions (Ba^{2+} : 13.7 eV (ref. 17); SO_4^{2-} : 9.8 eV derived from ref. 18). Starting the stimulation with the most energetically favourable nucleus of two Ba^{2+} and two SO_4^{2-} ions (outlined as the small rhombohedron in Fig. 3b), we first calculated the attachment energies for both species at their respective lattice positions around the step edges of the nucleus. Attachment energies of SO_4^{2-} ions at positions A and B in Fig. 3b are endothermic at +3 eV and +4 eV respectively, but are more favourable than attachment of Ba^{2+} to this nucleus. Allowing the SO_4^{2-} ions to relax away from these lattice positions reduces the attachment energy, and ions from both A and B relax to a position close to position A with an attachment energy of +0.9 eV, clearly demonstrating the anisotropy of the attachment process. As this value is still positive, starting a new row along the $[120]$ direction from A towards C is the rate-limiting step because the consecutive addition of further ions is exothermic and growth of each row continues towards position C. The same analysis was carried out after continued growth of the nucleus (A', B', C'). The energies for the attachment of Ba^{2+} ions to positions D and E, after adsorbing SO_4^{2-} to A', are comparable, and therefore growth of a new row can be started from E before the row through A' and D is complete. This phenomenon of growing rows with different lengths leads to the shape of the growth sector, which appears to be rounded towards the main growth direction (arrow labelled α in Fig. 3a, b). In the minor growth direction (β), the less

favourable addition of ions (and therefore a ten-fold slower growth) can only start at the corner position E, which keeps F angular during growth.

We have also calculated that the difference in energy between attaching an SO_4^{2-} ion to a Ba^{2+} ion on the (210) face in the $[1\bar{2}0]$ direction and in the $[\bar{1}20]$ direction is of the order of 1 eV. This explains the anisotropy of the rate for needle growth in these two directions on the (210) plane.

To model two-dimensional nuclei fully would involve an algorithm that dynamically models the statistical attachment of ions and includes the influence of solution composition and defects in the

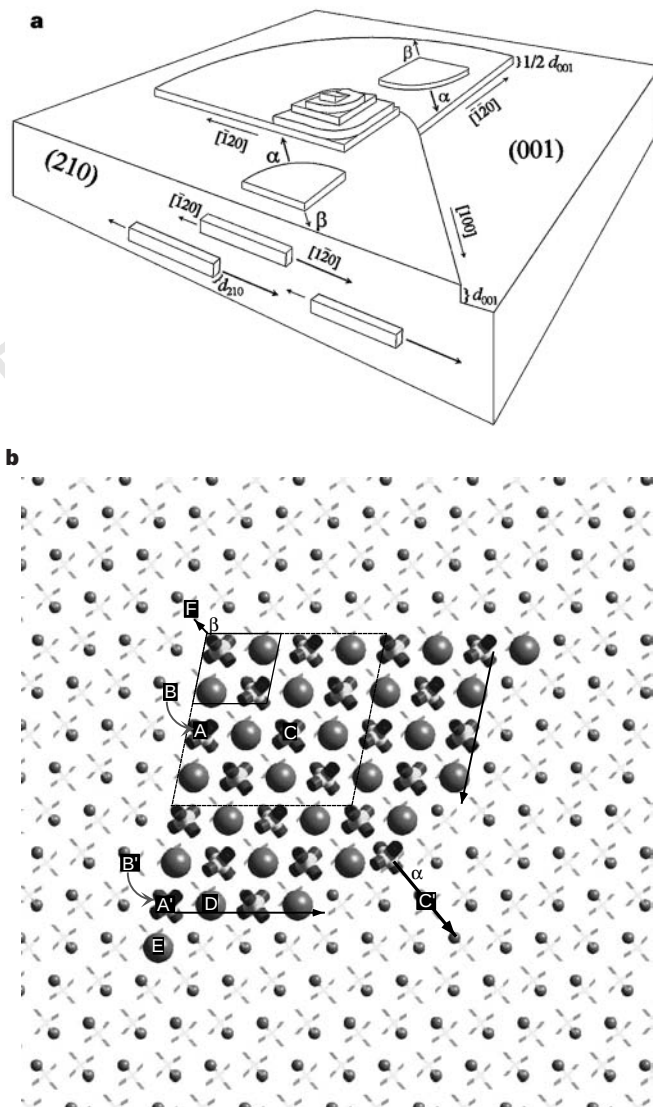


Figure 3 Illustration and computer simulation of crystal growth mechanisms. **a**, Diagram showing the crystallographic orientation of the two-dimensional nuclei on (001) and (210) surfaces, as shown in Figs 1 and 2. The lateral scales on each surface are greatly underestimated relative to the scale normal to the surface. **b**, Computer-simulated image of a sector-shaped, two-dimensional nucleus on a (001) surface. The Ba^{2+} positions and SO_4^{2-} groups are represented by spheres and tetrahedra respectively. The underlying barite substrate is shown in smaller symbols for clarity. The most energetically favourable nucleus (small solid parallelogram), positions of important attachment sites (A–F) and the growth directions (α , β) are indicated (see text). The microscopic growth directions along the dashed arrows lead to the mesoscopic growth direction α , as observed with AFM. In this image, all molecules outside the larger dashed parallelogram are allowed to relax. The positions A, A'; B, B' and so on are symmetry-equivalent in the bulk and define equivalent positions during growth (for example, A, A' are the starting positions for the growth of a new row).

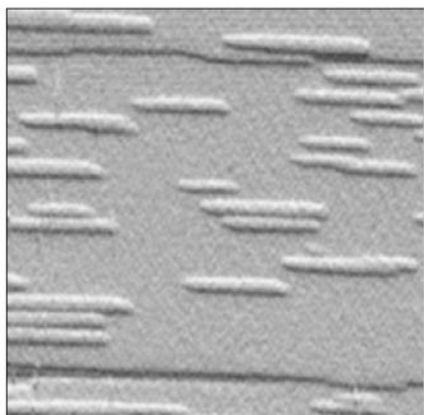


Figure 2 Two-dimensional nucleation on a (210) surface of barite. The nuclei are needle-shaped and grow along $\langle 120 \rangle$ directions. The fluid supersaturation value was 19; image area is $1.3 \mu\text{m} \times 1.3 \mu\text{m}$.

solid¹⁹. However, the deterministic approach presented here is sufficient to explain the observed anisotropic shape and growth kinetics of the nuclei.

The structural control and anisotropy of the growth explains the inhibition of spiral growth. On the (001) surface, alternate BaSO₄ layers are related by a 2₁ screw axis so that the anisotropy and the shape of the nuclei are reversed in each growth layer. The growth of the first BaSO₄ layer around the dislocation (Fig. 1b–d) is restricted to one sector, in the α direction (Fig. 3a), and cannot continue around the spiral because of the very slow growth in the opposite β direction. In the next layer that forms around the dislocation, the directions α and β are reversed but the rapid growth along α is almost completely prevented by the very slow growth in the underlying layer in this direction. This alternation of fast and slow directions continues for subsequent layers and so growth around the screw dislocation is limited to an ever-tightening spiral (Fig. 1f). Only when a half-layer from a neighbouring two-dimensional nucleus impinges on the spiral, as in the lower left-hand edge of the spiral in Fig. 1d, can the overlying layer spread laterally.

The {001} and {210} faces dominate the morphology of natural barite crystals, yet the anisotropy of the growth on these faces precludes spiral growth as a kinetically significant process. Only at higher supersaturation, when two-dimensional nucleation becomes a viable mechanism, can the spiral develop laterally. The anisotropy described here is not predicted by any crystal growth theories, principally because they do not take into account the actual surface structure and the attachment energies on specific surface sites. These attachment energies and hence the magnitude of the anisotropy could be modified by the presence of adsorbed impurities. Symmetry criteria allowing anisotropy exist in many crystal structures and therefore it is likely that these results will not only be valid for all compounds⁹ having the barite structure, but also more generally. This mechanism of inhibition of spiral growth may explain the “growth-stop phenomena” described in other systems⁴. Ultimately, the macroscopic crystal morphology depends on the microscopic processes of adsorption of growth units on crystal faces and step ledges and the methods used here provide powerful tools to observe directly and model these processes. □

Received 17 March; accepted 7 July 1998.

- Chernov, A. A. *Modern Crystallography III (Crystal Growth)* (Springer, Berlin, 1984).
- Sunagawa, I. in *Morphology of Crystals* (ed. Sunagawa, I.) 323–365 (Terra Scientific, Tokyo, 1987).
- Burton, W. K., Cabrera, N. & Frank, F. C. The growth of crystals and the equilibrium structures of their surfaces. *Phil. Trans. R. Soc.* **243**, 299–358 (1951).
- Bennema, P. Spiral growth and surface roughening: developments since Burton, Cabrera and Frank. *J. Cryst. Growth* **69**, 182–197 (1984).
- Bennema, P. & Gilmer, J. in *Crystal Growth: An Introduction* (ed. Hartman, P.) 263–327 (North Holland, Amsterdam, 1973).
- Botsaris, G. D. in *Industrial Crystallization 81* (eds Jancic, S. J. & de Jong, E. J.) (North Holland, Amsterdam, 1982).
- Nielsen, A. E. *Kinetics of Precipitation* (Pergamon, Oxford, 1964).
- Hartman, P. in *Morphology of Crystals* (ed. Sunagawa, I.) 269–319 (Terra Scientific, Tokyo, 1987).
- Hartman, P. & Strom, C. S. Structural morphology of crystals with the barite (BaSO₄) structure: a revision and extension. *J. Cryst. Growth* **97**, 502–512 (1989).
- Hillner, P. E., Gratz, A. J., Manne, S. & Hansma, P. K. Atomic-scale imaging of calcite growth and dissolution in real time. *Geology* **20**, 359–362 (1992).
- Dove, P. M. & Chermak, J. in *CMS Workshop Lectures: Scanning Probe Microscopy of Clay Minerals* (eds Nagy, K. L. & Blum, A. E.) 139–170 (Clay Minerals Soc., Washington DC, 1994).
- Allan, N. L. *et al.* Calculated bulk and surface properties of sulfates. *Faraday Disc.* **95**, 273–280 (1993).
- Miyake, M., Minato, I., Morikawa, H. & Iwai, S. Crystal structures and sulphate force constants of barite, celestine and anglesite. *Am. Mineral.* **63**, 506–510 (1978).
- Hartman, P. & Perdok, W. G. On the relations between structure and morphology of crystals. *Acta Crystallogr.* **8**, 48–52 (1955).
- Hottenhuis, M. H. J. & Lucasius, C. B. The influence of internal crystal structure on surface morphology; *in situ* observations of potassium hydrogen phthalate {010}. *J. Cryst. Growth* **94**, 708–720 (1989).
- Hartman, P. & Heijnen, W. M. M. Growth mechanism of a crystal face for which more than one surface structure is possible. *J. Cryst. Growth* **63**, 261–264 (1983).
- Conway, B. E. *Ionic Hydration in Chemistry and Biophysics* (Elsevier Scientific, Amsterdam, 1980).
- Rashin, A. A. & Honig, B. Reevaluation of the Born model of ion hydration. *J. Phys. Chem.* **89**, 5588–5593 (1985).
- Becker, U., Pina, C. M., Risthaus, P., Bosbach, D. & Putnis, A. Experimental and theoretical treatment of molecular-scale mechanisms of crystal growth on barite (Eighth annual V. M. Goldschmidt Conference, Toulouse, France, 1998).

Acknowledgements. We thank the Deutsche Forschungsgemeinschaft (DFG) and the Spanish Ministry of Science and Culture for financial support for this project, and A. Pina for Fig. 3a.

Correspondence and requests for materials should be addressed to A.P. (e-mail: putnis@mwz.uni-muenster.de).

Satellite mapping of enhanced BrO concentrations in the troposphere

T. Wagner & U. Platt

Institut für Umweltphysik, University of Heidelberg, INF 366, 69120 Heidelberg, Germany

Reactive bromine species contribute significantly to the destruction of ozone in the polar stratosphere¹. Reactive halogen compounds can have a strong effect not only on the chemistry of the stratosphere but also on that of the underlying troposphere. For example, severe ozone depletion events that are less persistent than those in the stratosphere occur in the Arctic² and Antarctic³ boundary layer during springtime and are also associated with enhanced BrO abundances^{2,4–10}. Observations^{5–8} of BrO (and ClO, which is less important) at ground level during these ozone depletion events have revealed halogen oxide mixing ratios of up to 30 parts per trillion—sufficient to destroy within one to two days the 30–40 parts per billion of ozone typically present in the boundary layer. The catalytic mechanism leading to so-called ‘tropospheric ozone holes’ is well established^{2,11}, but the origin of the increased BrO concentrations and the spatial and temporal extent of these events remains poorly understood. Here we

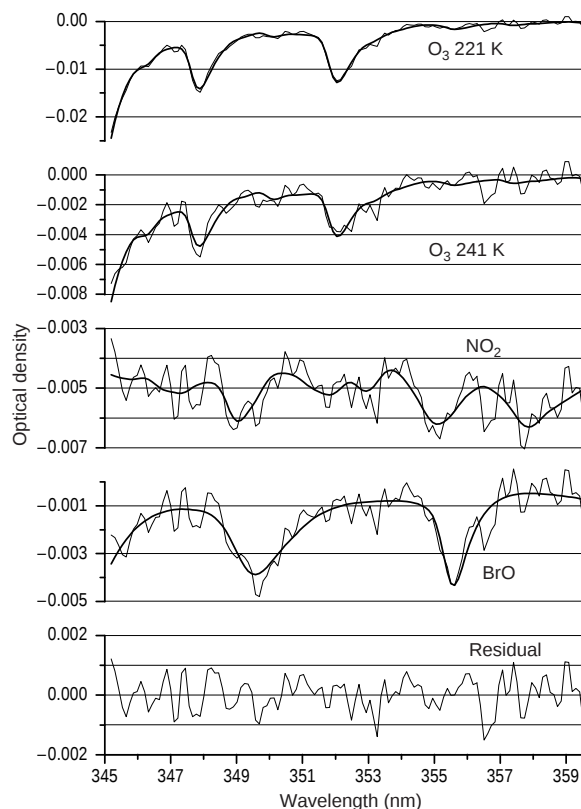


Figure 1 Results of the BrO evaluation procedure for a single spectrum taken on 15 September 1996, 22:30:09 (latitude: 66.2°S, longitude: 152.9°E, SZA: 76.9°). The reference spectra of the trace gases (thick lines) determined in the modelling process are compared with the respective absorption features in the measured spectrum (thin lines). The spectra of (O₂)₂ and OClO, which are also included in the fitting procedure, are not shown. The determined BrO SCD is 4.4×10^{14} mol cm⁻².

present satellite observations showing that tropospheric air masses enriched in BrO are always situated close to sea ice and typically extend over areas of about 300–2,000 km. The BrO abundances remain enhanced for periods of 1 to 3 days. These observations support the suggestion^{7,9,10} that autocatalytic release of bromine from sea salt gives rise to significant BrO formation which, in turn, initiates ozone depletion in the polar troposphere.

The GOME instrument is one of several instruments aboard the European research satellite ERS-2 (ref. 12). It consists of a set of four spectrometers that measure sunlight reflected from the Earth in the wavelength range 240–790 nm, and have moderate spectral resolution (0.2 nm/0.4 nm) and therefore enables the absorption of different atmospheric trace gases to be detected¹⁰. The satellite operates in a nearly polar, Sun-synchronous orbit of an altitude of 780 km, with a local Equator-crossing time of ~10:30. While the satellite moves in an almost north–south direction, the GOME instrument scans in an east–west direction. During each scan, three individual measurements are made, the corresponding ‘ground pixels’ (west, centre, and east pixel) each cover an area of 320 km (east–west) by 40 km (north–south). At the Equator, the Earth’s surface is covered in three days; polewards from ~70°, it is covered daily.

From the measured raw spectra of the GOME instrument, slant column densities (SCD, or the integrated concentration along the light path) of the absorbing trace gases are calculated¹³ by applying

differential optical absorption spectroscopy¹⁴ (DOAS). In brief, the measured spectra are modelled using suitably weighted trace-gas and sunlight reference spectra¹⁵ (Fig. 1). From the model parameters and the trace gas absorption cross-sections (for BrO, the data are from ref. 16), the desired trace-gas SCDs are calculated. The light reaching the instrument is either reflected from the Earth’s surface or scattered back from the atmosphere. For that reason, the determination of the vertical column density (VCD, the vertically integrated trace-gas concentration) from the measured SCD requires radiative transport modelling to derive air mass factors (AMF), where $AMF = SCD/VCD$. We calculated AMFs using a Monte Carlo radiative transport model including spherical geometry and multiple scattering¹⁷. Two sets of AMFs were used: one for the stratospheric part (assuming maximum BrO concentration at 17,000 m altitude) and one for the tropospheric part of the BrO column (assuming a constant concentration between 0 m and 1,000 m altitudes, where O₃ depletion was typically observed in the Arctic³ and Antarctic⁴). As clouds reflect light and thereby ‘hide’ trace gases in the part of the atmosphere that is below them, it is important to know whether the atmosphere was (at least partly) cloud free during the observations.

The state of cloudiness can be quantified by monitoring the absorption due to atmospheric (O₂)₂, the O₂–O₂ collision complex. The total atmospheric column density of (O₂)₂ varies only slightly with pressure and temperature. More important, most (O₂)₂ is

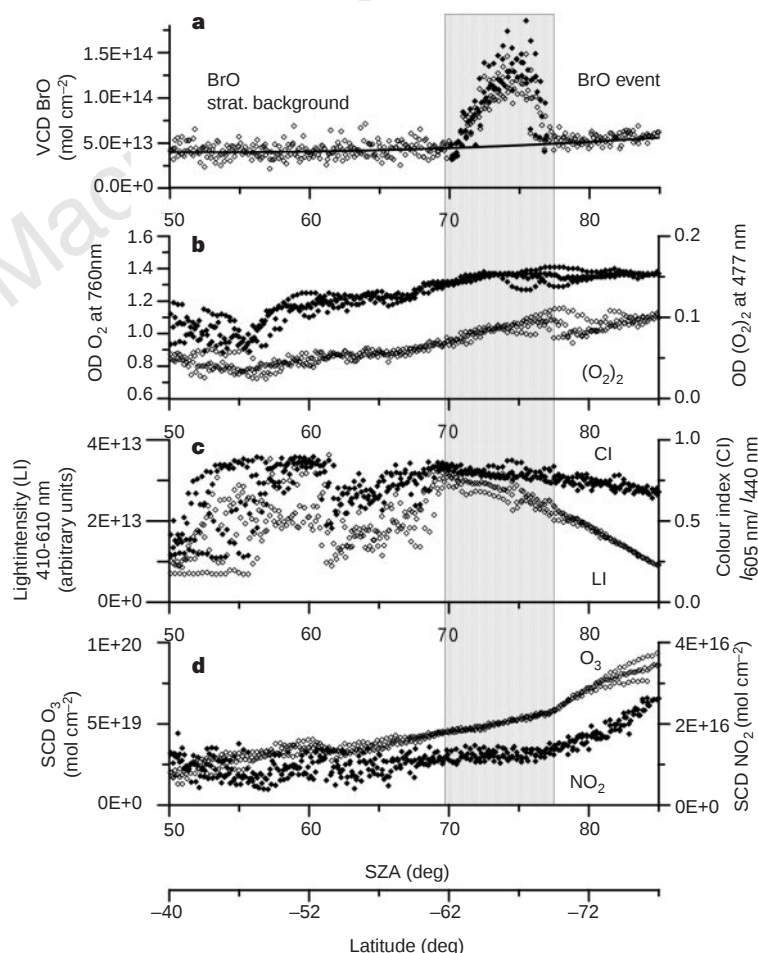


Figure 2 Latitudinal variation of the column densities of atmospheric trace gases, the average light intensity and a colour index for part of an orbit on 15 September 1996 (ERS-2 orbit 60915214) as a function of the solar zenith angle and corresponding (approximate) latitude. **a**, Open diamonds, BrO VCD calculated using the stratospheric AMF; filled diamonds, the enhancement expressed as tropospheric BrO VCD (calculated using tropospheric AMFs as described in the

text). **b**, Open diamonds, absorption of (O₂)₂ (right axis); filled diamonds, absorption of O₂ (left axis). **c**, Open diamonds, measured average intensity (410–610 nm, left axis); filled diamonds, colour index (intensity ratio at the wavelengths 605 and 440 nm, right axis). **d**, Open diamonds, SCD of O₃ (left axis); filled diamonds, SCD of NO₂ (right axis).

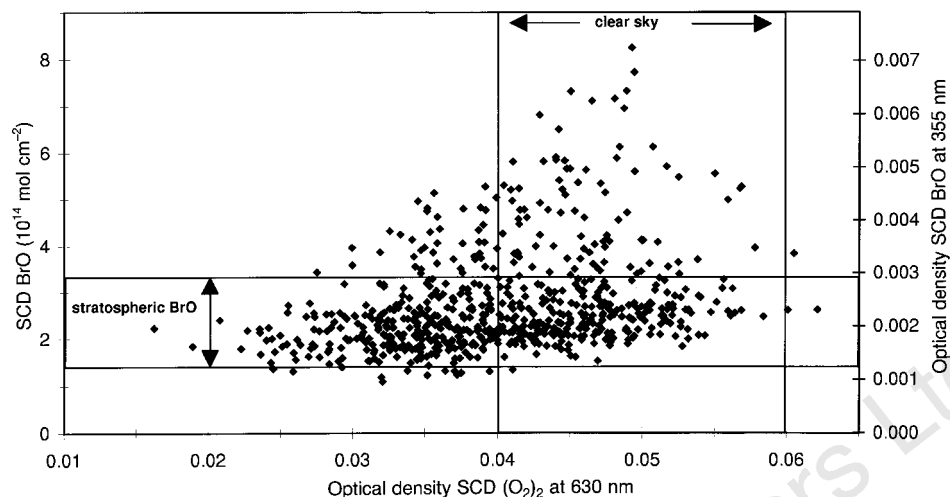


Figure 3 Scatter plot of simultaneously measured BrO and $(\text{O}_2)_2$ SCDs (optical densities) for 2 to 15 September 1996. The vertical box indicates the range of $(\text{O}_2)_2$ absorptions for clear sky conditions; the horizontal box indicates the range of

purely stratospheric BrO slant column densities. In the chosen latitude range between 63°S and 67°S (SLA between 71° and 79°), several events of enhanced BrO absorption were detected.

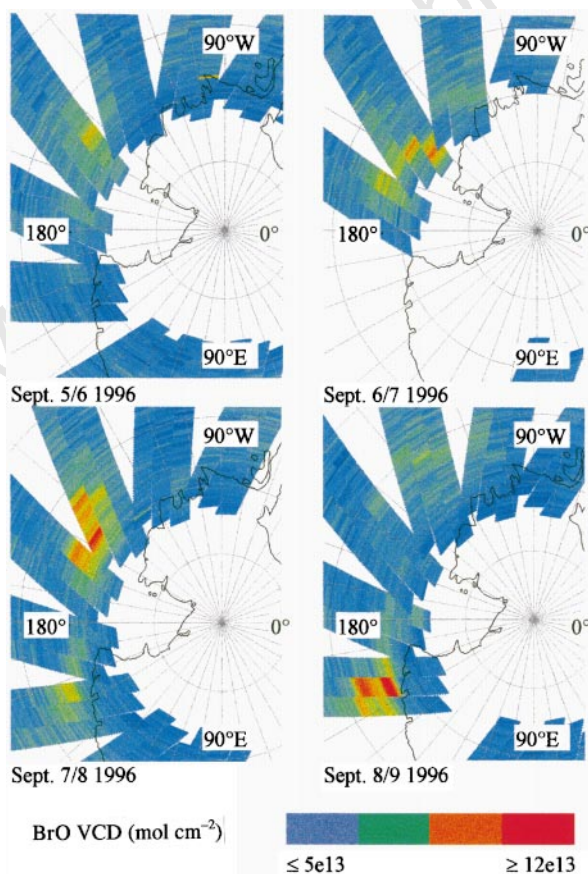


Figure 4 GOME observations of the BrO vertical column density over the Arctic from 4/5 to 8/9 September 1996. The areas where elevated tropospheric BrO

concentrations build up and decay are clearly visible around the Antarctic continent (no satellite data are available for the white areas on the map).

located below the cloud cover in the lower troposphere (with a scale height of $\sim 4,000\text{ m}$)¹⁸. The GOME instrument will therefore detect the usually high atmospheric $(\text{O}_2)_2$ absorption only under (at least partly) clear sky conditions, thus making $(\text{O}_2)_2$ absorption measurements a sensitive indicator of cloudiness.

The tropospheric BrO VCDs reported here are associated with a relative error of $\sim 30\%$. This value reflects uncertainties in the extent of cloud cover, an uncertainty of about 8% in the cross-section of BrO (ref. 16), an uncertainty of $\sim 10\%$ in calculated AMFs (derived

from using varying atmospheric parameters in the radiative transfer model) and a relative error in the modelled BrO absorption of $\sim 9\%$. (The error associated with BrO absorption was calculated using the analysis of Stutz and Platt¹⁵; the effects associated with a change in the instrument's resolution between spectra from direct sunlight and earthshine (K. Chance, personal communication) were found to be negligible. Further details of error analysis and uncertainty estimates are provided as Supplementary Information.) Spatial averaging within a GOME ground pixel can give rise to additional

Table 1 Summary of events of enhanced tropospheric BrO concentrations

Date of start	Approx. duration (days)	Location (centre)	Max. BrO concentration* (10^8 mol cm^{-3})	Extension north-south (km)	Extension east-west (km)
5 Sept. 1996	3	150° W, 67° S	14.9	1,000	1,400
7 Sept. 1996	3	155° E, 67° S	12.3	600	700
11 Sept. 1996	2	180° W, 69° S	6.8	400	1,400
14 Sept. 1996	2	160° E, 66° S	11.1	700	2,000
16 Sept. 1996	1	177° E, 68° S	6.9	300	300
17 Sept. 1996	3	160° W, 74° S	11.9	1,200	700
20 Apr. 1997	2	30° W, 79° N	5.0	700	600

* If a BrO layer height of 1,000 m is assumed.

Enhanced tropospheric BrO concentrations observed in Antarctica (from 90° E to 90° W and from 60° S to 75° S, where enhanced BrO absorptions were observed frequently) between 4 and 19 September. Also shown are data of one observation in the Arctic on 20 April 1997, when simultaneously low O_3 concentrations were measured at Ny Ålesund (Spitsbergen). The duration of the events was estimated from measurements with VCDs to be at least 30% higher than for purely stratospheric levels.

errors as local BrO maxima may exceed the average values reported here.

Figure 2 shows data derived from measurements obtained during the southern part of one satellite orbit, when enhanced BrO absorption was detected near 67° S, 161° E. For solar zenith angles below 70° or above 78° (corresponding to latitudes 62° S and 70° S, respectively), the VCD values are typical for stratospheric BrO, but rise by a factor of 2 to 3 when the satellite covers the area in between (Fig. 2a). We attribute this enhancement to tropospheric BrO. The difference between VCDs in this region and the purely stratospheric VCDs observed outside this region is used to calculate tropospheric BrO VCDs ($\text{VCD}_{\text{trop}} = (\text{VCD} - \text{VCD}_{\text{strat}}) \times \text{AMF}_{\text{strat}}/\text{AMF}_{\text{trop}}$). Figure 2 also shows the absorption due to O_2 and $(\text{O}_2)_2$ (Fig. 2b) and the measured light intensity parameters (Fig. 2c), which are indicators of changes in the radiative properties of the atmosphere (as caused by clouds, for example). These four parameters are fairly constant throughout the period of enhanced BrO, as are the SCDs of the mainly stratospheric absorbers O_3 and NO_2 (Fig. 2d), indicating that the increase in BrO absorption is most probably caused by an increased abundance of tropospheric BrO, rather than by a disturbance of the stratospheric composition or a modification of the stratospheric AMF. This conclusion is corroborated by the results in Fig. 3, which shows BrO SCDs as a function of the simultaneously observed $(\text{O}_2)_2$ SCDs¹⁹: under both clear and cloudy sky conditions, BrO densities have been observed, within the range of typical stratospheric values, whereas significantly enhanced BrO densities were only detected under (at least partly) clear sky conditions, when the spectrometers aboard the satellite were able to 'look' deep down into the troposphere.

Typical events of enhanced BrO absorption are shown in Fig. 4, which illustrates the formation and decay of 'BrO absorption patches' over a period of four days in September 1996. Table 1 summarizes the events between 4 and 19 September 1996. It also includes an event observed on 20 April 1997 over Spitsbergen, which occurred while ground-based measurements at Ny Ålesund (Spitsbergen) indicated a depletion in tropospheric O_3 (E. Lehrer, private communication). As indicated in Table 1, the tropospheric air masses enriched in BrO extend over large areas (average value is $73,000 \text{ km}^2$), with BrO concentrations reaching maximum values of $5\text{--}15 \times 10^8 \text{ mol cm}^{-3}$. These values correspond to BrO mixing ratios of 20–60 p.p.t. and thus agree with the results of ground-based measurements during "polar tropospheric ozone hole" events⁶. Note, however, that the BrO concentrations listed in Table 1 have been calculated assuming a planetary boundary layer height of 1,000 m: although this value provides a good estimate of the average thickness of the boundary layer over Antarctica^{4,20}, observed thicknesses range from about a few hundred metres to ~2 km (refs. 3, 4) and result in even higher or lower BrO concentrations derived from the observed VCD values.

The enhanced BrO absorption events that have been detected in the GOME data on several occasions are difficult to explain other than by assuming an increase in tropospheric BrO. In particular, the observed BrO VCDs far exceed the stratospheric BrO column densities predicted by models (or, in fact, even the total inorganic

Br column densities in the stratosphere). Furthermore, the events that we have observed all took place during springtime and over polar ice sheets; that is, at times and locations typical for the occurrence of enhanced tropospheric BrO associated with "tropospheric ozone holes". The mechanism leading to the formation of these ozone holes, that is, to ozone destruction in the troposphere, is simple and involves the oxidation of Br or Cl to the corresponding halogen oxide, which subsequently reacts with BrO to regenerate two halogen atoms^{2,11}. This catalytic cycle is terminated once Br and Cl can no longer be regenerated, probably because of the formation of much less reactive HBr and organobromine compounds.

The origin and formation of the enhanced abundance of BrO and Br during tropospheric ozone destruction events are not yet fully understood. The observations reported here provide valuable information in this context as satellite data, in contrast to ground-based observations, allow for a clear distinction between spatial and temporal variations in BrO enhancement in the troposphere. The observed duration of 1 to 3 days suggests that the high concentrations of BrO (and hence of Br) in the troposphere are due to autocatalytic Br release⁷, probably from sea salt deposits on sea ice^{7,9,10}, rather than to the degradation of unstable organic compounds containing halogens². We anticipate that the detailed study of further observations may reveal how the BrO formation process propagates at the edges of processed air masses, thus providing addition insight into the destruction of tropospheric ozone in polar regions. We also consider that global satellite observations will contribute to a more comprehensive assessment of the potential importance of BrO for tropospheric chemistry more generally. □

Received 18 November 1997; accepted 7 July 1998.

- McElroy, M., Salawitch, R., Wofsy, S. & Logan, J. Reductions of Antarctic ozone due to synergistic interactions of chlorine and bromine. *Nature* **321**, 795–762 (1986).
- Barrie, L. A., Bottenheim, J. W., Schnell, R. C., Crutzen, P. J. & Rasmussen, R. A. Ozone destruction and photochemical reactions at polar sunrise in the lower Arctic atmosphere. *Nature* **334**, 138–141 (1988).
- Solberg, S., Schmidbauer, N., Semb, A., Stordal, F. & Øystein, H. Boundary layer ozone depletion as seen in the Norwegian Arctic in spring. *J. Atmos. Chem.* **23**, 301–332 (1996).
- Wessel, S. *Troposphärische Ozonvariationen in Polarregionen* (Thesis, Univ. Bremen, 1996).
- Hausmann, M. & Platt, U. Spectroscopic measurement of bromine oxide and ozone in the high Arctic during Polar Sunrise Experiment 1992. *J. Geophys. Res.* **99**, 25399–25414 (1994).
- Tuckermann, M. et al. DOAS observation of halogen radical-catalysed Arctic boundary layer ozone destruction during the ARCTOC-campaigns 1995 and 1996 in Ny Ålesund, Spitsbergen. *Tellus* **49 B**, 533–555 (1997).
- Platt, U. & Lehrer, E. ARCTOC final report to EU, Brussels (1996).
- Kreher, K., Johnston, P. V., Wood, S. W. & Platt, U. Ground-based measurements of tropospheric and stratospheric BrO at Arrival Heights (78° S), Antarctica. *Geophys. Res. Lett.* **24**, 3021–3024 (1997).
- Tang, T. & McConnell, J. C. Autocatalytic release of bromine from arctic snow pack during polar sunrise. *Geophys. Res. Lett.* **23**, 2633–2636 (1996).
- Vogt, R., Crutzen, P. J. & Sander, R. A mechanism for halogen release from sea-salt aerosol in the remote marine boundary layer. *Nature* **383**, 327–330 (1996).
- Le Bras, G. & Platt, U. A possible mechanism for combined chlorine and bromine catalyzed destruction of tropospheric ozone in the Arctic. *J. Geophys. Res.* **22**, 599–602 (1995).
- ESA Publication Division GOME, *Global Ozone Monitoring Experiment, Users Manual*. (ed. Floyed, B.) (European Space Research and Technology Centre (ESTEC), Noordwijk, 1995).
- Hegels, E. et al. in *Atmospheric Ozone, Proc. 18th Quadrennial Ozone Symp.* (eds Bojkov, R. D. & Visconti, G.) 293–296 (L'Aquila, 1998).
- Platt, U. in *Air Monitoring by Spectroscopic Techniques* (ed. Sigrist, M. W.) 127 (Chemical Analysis Ser., Wiley, New York, 1994).
- Stutz, J. & Platt, U. Numerical analysis and error estimation of differential optical absorption spectroscopy measurements least-squares methods. *Appl. Optics* **35**, 6041–6053 (1996).
- Wahner, A., Ravishankara, A. R., Sander, S. P. & Friedl, R. R. Absorption cross-section of BrO between 312 and 385 nm at 298 and 223 K. *Chem. Phys. Lett.* **152**, 507–512 (1988).
- Marquard, L. C. & Platt, U. AMFTRAN: A new Monte Carlo radiative transfer model for the

- calculation of air mass factors. *Proc. NATO ARW, Athens, Nov. 1995*. (Springer, Heidelberg, London, New York, 1997).
18. Osterkamp, H. et al. in *Air Pollution Research Report 66, Polar Stratospheric Ozone 1997, Proc. 4th European Symp.* (eds Harris, N. R. P., Kilbane-Dawe, I. & Amanatidis, G. T.) 478–481 (Office for Official Publications of the European Communities, Luxembourg, 1998).
 19. Wagner, T. et al. in *Air Pollution Research Report 66, Polar Stratospheric Ozone 1997, Proc. 4th European Symp.* (eds Harris, N. R. P., Kilbane-Dawe, I. & Amanatidis, G. T.) 514–517 (Office for Official Publications of the European Communities, Luxembourg, 1998).
 20. Fortuin, J. P. F. & Orlemans, J. An Atmospheric Model for Simulating the Mass Balance and Temperature on the Antarctic Ice Sheet. *Zeitschrift für Gletscherkunde und Glazialgeologie* **26**, 31–56 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. The financial support of the Deutsche Agentur für Raumfahrtangelegenheiten (DARA) is acknowledged. We thank the ESA operation centre in Frascati (Italy) and the DLR, Department 'Wissenschaftlich-technische Betriebseinrichtungen' in Wessling for supplying the satellite data; C. Leue for developing the computer graphics program; L. Marquard for calculating AMFs; and A. Richter for discussion.

Correspondence should be addressed to T.W.

Hydrothermal activity along the southwest Indian ridge

C. R. German*, E. T. Baker†, C. Mevel‡, K. Tamaki§ & the FUJI Science Team||

* Southampton Oceanography Centre, Empress Dock, Southampton SO14 3ZH, UK

† NOAA-Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115-0070, USA

‡ CNRS-Laboratoire de Petrologie, Université Pierre et Marie Curie, Case 110, 4 place Jussieu, 75252 Paris Cedex 05, France

§ Ocean Research Institute, University of Tokyo, 1-15-1 Minamidai, Nakano-ku, Tokyo-164, Japan

Twenty years after the discovery of sea-floor hot springs, vast stretches of the global mid-ocean-ridge system remain unexplored for hydrothermal venting. The southwest Indian ridge is a particularly intriguing region, as it is both the slowest-spreading of the main ridges¹ and the sole modern migration pathway between the diverse vent fauna of the Atlantic and Pacific oceans². A recent model postulates that a linear relation exists between vent frequency and spreading rate³ and predicts vent fields to be scarcest along the slowest-spreading ridge sections, thus impeding migration and enhancing faunal diversity². Here, however, we report evidence of hydrothermal plumes at six locations within two 200-km-long sections of the southwest Indian ridge indicating a higher frequency of venting than expected. These results suggest that fluxes of heat and chemicals from slow-spreading ridges may be greater than previously thought and that faunal migration along the southwest Indian ridge may serve as an important corridor for gene-flow between Pacific and Atlantic hydrothermal fields.

Submarine hydrothermal springs at mid-ocean ridges emit hot, anoxic fluids which mix turbulently with ambient sea water as they rise from the sea floor, precipitating various sulphide and oxide phases, until some level of neutral buoyancy is attained hundreds of metres above the sea floor^{4,5}. Resultant plumes are then dispersed along isopycnal surfaces where they can be readily detected from their strong enrichment in dissolved chemical tracers and increased suspended particulate load⁶. In a recent study, we took advantage of

these characteristics to prospect for and identify evidence of hydrothermal venting along the very-slow-spreading southwest Indian ridge (SWIR). The SWIR separates the African and Antarctic plates, and spreads¹ at a full rate of 14 mm yr⁻¹. This very-slow-spreading ridge is also associated with cold mantle temperatures and thick lithosphere, and represents an end-member in the global mid-ocean-ridge system⁷. Using an approach developed to detect hydrothermal plumes along the Mid-Atlantic Ridge⁸, we attached optical back-scatter instruments to the cable of a deep-tow sidescan sonar instrument as it conducted geophysical surveys of the SWIR rift-valley floor in two areas⁹; these were between 58° 25' E and 60° 15' E, and between 63° 20' E and 65° 45' E (Fig. 1). The TOBI (towed ocean bottom instrument) sidescan was towed along two parallel tow-tracks, offset by 5–6 km, through each 200-km section of the SWIR rift-valley; the vehicle was towed at a speed of 2 knots and at heights ranging between 100 m and 500 m off-bottom. A suite of six self-logging MAPR (miniature autonomous plume recorder) instruments¹⁰ were placed at 50-m intervals from 100 m below to 200 m above the vehicle, thereby effecting the investigation of a 300-m-thick vertical swath of the water column overlying the tow-tracks.

Following this approach, six different sets of plume signals were detected (Fig. 2). The positions at which the maximum plume signals in each area were observed, together with the strengths of those signals and the plume depths at which they were recorded, are listed in Table 1. Each plume was identified by more than one sensor along the same track and in all cases, layers of particle-rich water were bounded, above and below, by the optically clear water more typical of the deep ocean far from land. This is an important observation because neutrally buoyant plumes from high-temperature hydrothermal vents are the only known source for such features above mid-ocean ridges. Sediment resuspension—the only other plausible source of suspended material at these depths—would be expected to cause continuous particle enrichments into the seabed, with increasing intensity downwards, rather than giving rise to the discrete lens-like layers of midwater particle-enrichment reported here. The fact that the plumes reported here are trapped within the SWIR rift-valley suggests that they are the product of chronic venting rather than “event” plumes as discussed recently by Palmer and Ernst¹⁸.

Three of the six plumes were identified from both parallel, but offset, tow-tracks of the TOBI vehicle. This is clearly exemplified by the strongest of the plumes identified during our survey: plume 2 (Table 1, Fig. 2). This plume can be readily identified from

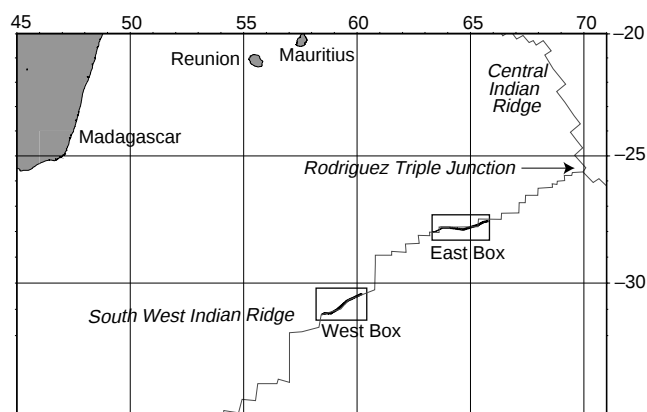


Figure 1 Regional map of the southwest Indian ridge (SWIR) spreading centre. Shown are two study areas investigated during the FUJI (French-UK-Japanese InterRidge) cruise aboard RV *Marion Dufresne*, October 1997. Also shown are the southern extent of the central Indian ridge and the Rodriguez triple junction. TOBI/MAPR study areas expanded in Fig. 2 are indicated here as the “West Box” (58° 10' to 60° 20' E) and the “East Box” (63° 20' to 65° 50' E).

|| P. Blondel, A. Evans, C. Flewellen, D. Matthew, L. Parson & A. Wallace (Southampton Oceanography Centre, Southampton, UK); D. Boulanger, D. Sauter (Ecole et Observatoire de Physique du Globe de Strasbourg, Strasbourg, France); H. Fujimoto, Y. Osada, C. Tamura, T. Yama-ashi (Ocean Research Institute, University of Tokyo, Japan); O. Gomez (Laboratoire d'Etudes en Geophysique et Oceanographie Spatiales, Toulouse, France); S. M. Lee, R. C. Searle (Dept of Geological Sciences, Durham, UK); P. Patriat, G. Poullequin (Institut de Physique de Globe de Paris, Paris, France); C. Rommevaux-Jestin (CNRS-Université Pierre et Marie Curie, Paris, France); T. Yamada (Earthquake Research Institute, Tokyo, Japan)

cross-sections of optical backscattering data, centred at a water depth of 4,100–4,200 m (Fig. 3). Along southwest–northeast section AA', strongest plume anomalies were observed between 2 and 5 km along-track. It is important to note the undulations in the MAPR track-paths as the TOBI vehicle was first lowered and then raised through these water depths. No particle-rich anomalies are apparent further east along AA', at 7–8 km along track, where the TOBI vehicle was lowered back through the same water depths. This provides a strong indication that the source of venting was most closely approached between 4 and 5 km along this track, near 31° 06' E, 59° 01' S (Fig. 2). During the return pass through the same segment, even higher maximum plume anomalies were observed at the same plume depths (section BB', Fig. 3). Maximum anomalies along this second section occurred between 3 and 4 km along-track at 31° 04' S, 58° 58' E (Table 1). The light-scattering anomaly at this location was equivalent to a suspended particulate load of $\sim 20 \mu\text{g l}^{-1}$ based on prior field calibration of the back-scattering sensors¹¹. For plume 2, we note that the neutrally buoyant plume occurs at a depth of $>4,000$ m. Therefore, its source of venting must also occur below 4,000 m depth. This represents a greater depth than any previously identified sea-floor hydrothermal field¹².

The results reported here represent, to our knowledge, the first evidence that high-temperature hydrothermal activity occurs along the very-slow-spreading SWIR. Extinct hydrothermal sulphide deposits have been recovered from the central Indian ridge near the Rodriguez triple junction¹³, and plumes rich with Mn, CH₄ and particles have also been detected between 23° S and 25° S along the central Indian ridge close to the triple junction^{13,14}. Evidence for hydrothermal plumes overlying the medium–fast-spreading southeast Indian ridge (SEIR) have also been reported recently¹⁵. Along the SEIR, however, full spreading rates are $\sim 65 \text{ mm yr}^{-1}$ and so relatively abundant venting had been expected, comparable to the incidence of venting observed on the Juan de Fuca ridge (spreading rate 55 mm yr^{-1})³. By contrast, because of the very-slow spreading rate of the SWIR (14 mm yr^{-1}), venting had been expected at no more than two-thirds of the frequency detected previously along the slow-spreading Mid-Atlantic Ridge^{6,16} suggesting a plume incidence of no more than one vent site every 200–300 km along axis. Instead, we have found evidence for at least six discrete sources of hydrothermal plume discharge at an average spacing of <100 km. This lends further support to recent suggestions that tectonic processes can increase the abundance of hydrothermal venting along slow-spreading

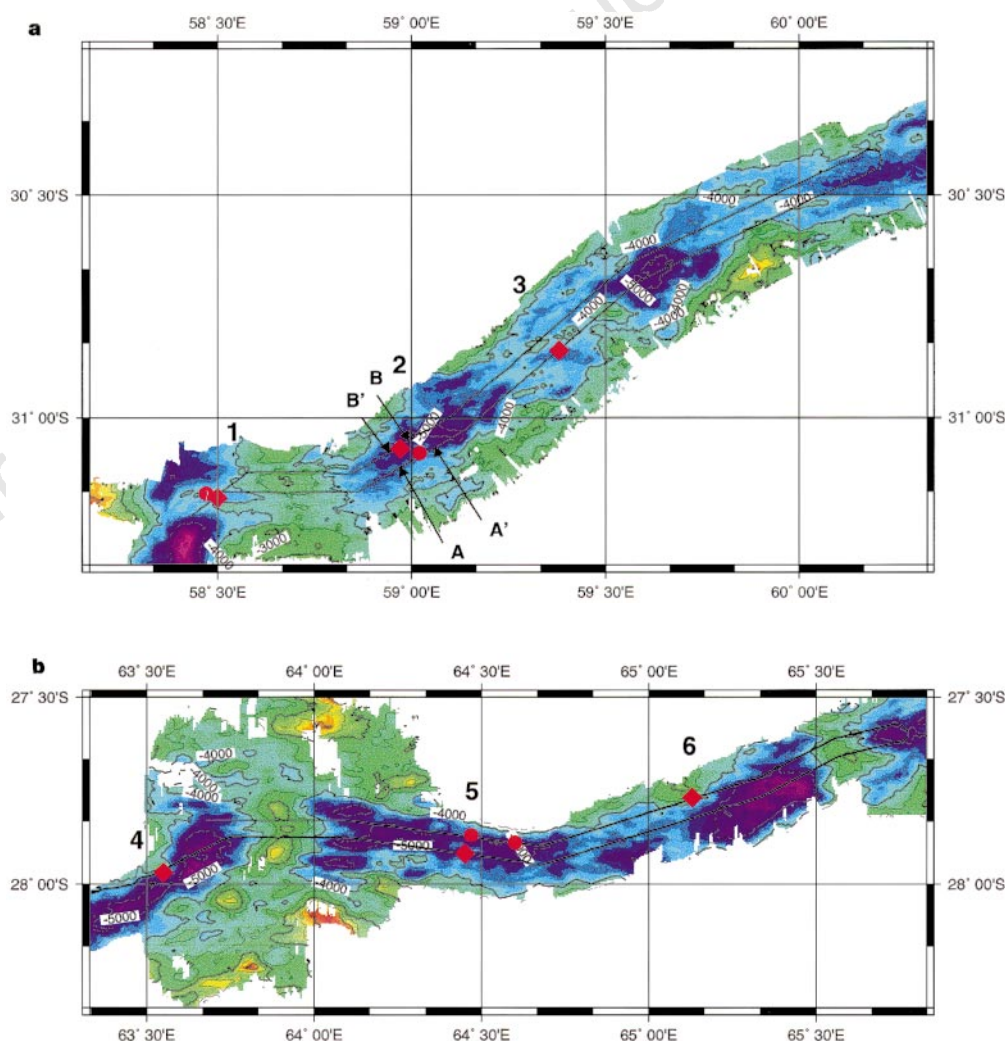


Figure 2 Bathymetric maps of the SWIR study areas. **a**, Western box between 58° 10' E and 60° 20' E; **b**, eastern box between 63° 20' E and 65° 50' E. Solid along-axis lines show offset outward (west-east) and return (east-west) TOBI tow-tracks within the deep rift-valley of the SWIR. Plume areas numbered 1–3 (West box) and 4–6 (East box) correspond to discrete areas at which evidence for

venting was observed. Red diamonds indicate locations of maximum plume anomalies listed in Table 1; red circles indicate locations of secondary plume maxima observed along parallel tow-tracks (at the same plume height but with weaker intensities) for plumes 1, 2 and 5. Also shown are the locations of MAPR cross-sections AA' and BB' (Fig. 3).

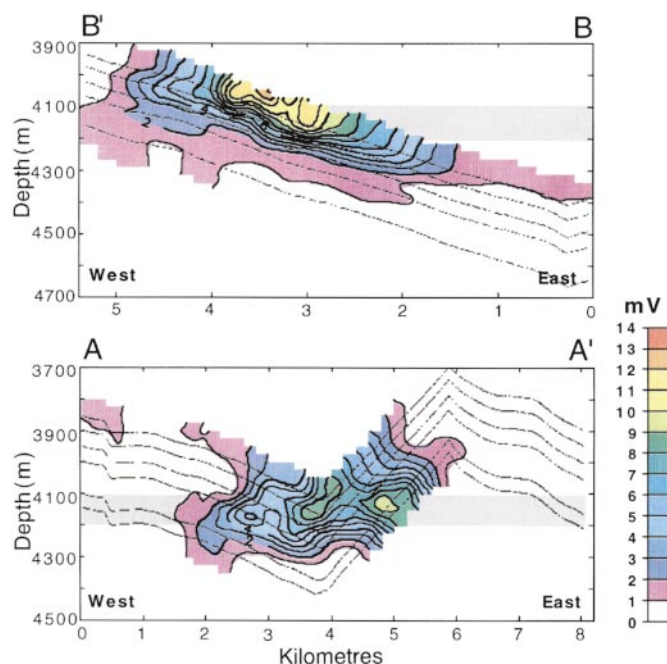


Figure 3 Cross-sections of plume 2, centred at 4,100–4,200 m water depth (grey shaded band), as detected from MAPRs arrayed along the tow cable above and below the side-scan vehicle TOBI. Section AA' runs southwest to northeast between 31°06.5' S, 58°58.5' E and 31°04' S, 59°03.5' E; section BB' runs north-east to southwest between 31°03' S, 58°59.5' E and 31°05' S, 59°56.5' E (Fig. 2). Units on colour scale are light back-scattering intensities in millivolts. Relative

back-scattering anomalies were calculated by first detrending the optical records to remove a slight voltage drift (typically <5 mV over a 5 day tow). Each optical record was then normalized to zero millivolts at depths of minimum back-scatter (3,300–3,400 m, 4,800–5,000 m) to account for slight differences between individual sensors. Maximum plume intensity in BB' is roughly $20 \mu\text{g l}^{-1}$, based on prior field calibration of the back-scattering sensors¹¹.

ridges beyond that which would be predicted from spreading rate alone⁸.

The identification of abundant venting along a very-slow-spreading mid-ocean ridge has significant implications for marine geochemistry and biology. First, current global estimates of thermal and chemical flux through high-temperature hydrothermal circulation disagree by at least one order of magnitude¹⁷. The demonstration that venting occurs relatively frequently along one of the very slowest-spreading sections of the global mid-ocean-ridge crest may, therefore, indicate that global fluxes are towards the higher end of present estimations. Second, if the distribution of venting along ridge-axes is not a simple function of spreading rate (magma supply), then chemical fluxes from hydrothermal venting to the deep ocean may not be concentrated in the deep Pacific Ocean to the same extent as has previously been expected. Instead, fluxes to the Atlantic and Indian oceans may both be greater than expected, with important implications for the deep-ocean geochemical cycles of various tracers. Last, vent ecosystems in the Pacific and Atlantic oceans host quite different biological communities, with closest convergence in the southwest Pacific². For the past 95 Myr, assuming that colonization of vent-specific organisms occurs through migration along-axis, the only viable communication route between these communities has lain along the SEIR and SWIR, into the southern

Atlantic Ocean². Our study has shown that the environmental niches required to support vent-specific organisms should be relatively abundant along the SWIR. The future location, examination and identification of the organisms inhabiting these unique vent sites, some in water depths >4,000 m, would represent a significant advance in our understanding of vent biogeography. □

Received 27 February; accepted 26 June 1998.

- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions *Geophys. J. Int.* **101**, 425–478 (1990).
- Tunnicliffe, V. & Fowler, C. M. R. Influence of seafloor spreading on the global hydrothermal vent fauna. *Nature* **379**, 531–533 (1996).
- Baker, E. T., Chen, Y. J. & Phipps Morgan, J. The relationship between near-axis hydrothermal cooling and the spreading rate of mid-ocean ridges. *Earth Planet. Sci. Lett.* **142**, 137–145 (1996).
- Lupton, J. E., Delaney, J. R., Johnson, H. P. & Tivey, M. K. Entrainment and vertical transport of deep ocean water by buoyant hydrothermal plumes. *Nature* **316**, 621–623 (1985).
- Feely, R. A. *et al.* Composition and dissolution of black smoker particulates from active vents on the Juan de Fuca Ridge. *J. Geophys. Res.* **92**, 11347–11365 (1987).
- Baker, E. T., German, C. R. & Elderfield, H. *Hydrothermal Plumes over Spreading-center Axes: Global Distributions and Geophysical Inferences* 47–71 (AGU Monogr. 91, Am. Geophys. Union, Washington DC, 1995).
- Langmuir, C. H. *et al.* *Investigation of the Southwest Indian Ridge: A Project Plan* (InterRidge SWIR Project Plan, InterRidge, Paris, 1997).
- German, C. R., Parson, L. M. & the HEAT Scientific Team. Hydrothermal exploration near the Azores triple-junction: tectonic control of venting at slow-spreading ridges? *Earth Planet. Sci. Lett.* **138**, 93–10 (1996).
- Tamaki, K. & Mevel, C. *FUJI (French–UK–Japan InterRidge) Cruise: R/V Marion Dufresne, Durban—la Reunion* (ORI Cruise Rep., Ocean Res. Inst., Tokyo, 1997).
- Baker, E. T. & Milburn, H. B. MAPR: A new instrument for hydrothermal plume mapping. *Ridge Events* **8**, 23–25 (1997).
- Feely, R. A. *et al.* Hydrothermal plume particles and dissolved phosphate over the superfast spreading southern East Pacific Rise. *Geochim. Cosmochim. Acta* **60**, 2297–2323 (1996).
- Rona, P. A. & Scott, S. D. A special issue on sea-floor hydrothermal mineralization: new perspectives. *Econ. Geol.* **88**, 1935–1976 (1993).
- Plüger, W. L. *et al.* Discovery of hydrothermal fields at the Central Indian Ridge. *Mar. Mining* **9**, 73–86 (1990).
- Gamo, T. Hydrothermal plumes at the Rodriguez triple junction, Indian Ridge. *Earth Planet. Sci. Lett.* **142**, 261–270 (1996).
- Scheirer, D. S., Baker, E. T. & Johnson, K. T. M. Detection of hydrothermal plumes along the Southeast Indian Ridge near the Amsterdam–St. Paul Plateau. *Geophys. Res. Lett.* **25**, 97–100 (1998).
- German, C. R., Baker, E. T. & Klinkhammer, G. *Regional Setting of Hydrothermal Activity* (Spec. Publ. 87, Geol. Soc., London, 1995).
- Elderfield, H. & Schulz, A. Mid-Ocean Ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**, 191–224 (1996).

Table 1 Hydrothermal plumes above the southwest Indian ridge

Plume	Maximum back-scatter anomaly* (mV)	Latitude, Longitude	Depth (m)
1	6.0	31°11' S, 58°30' E	3,500
2	17.0	31°04' S, 58°58' E	4,100
3	2.6	30°51' S, 59°23' E	3,800
4	3.8	27°58' S, 63°33' E	3,300
5	3.9	27°55' S, 64°27' E	3,900
6	1.6	27°46' S, 65°08' E	3,100

* See Fig. 3 legend for information on this quantity.

18. Palmer, M. R. & Ernst, G. G. Generation of hydrothermal megaplumes by cooling of pillow basalts at mid-ocean ridges. *Nature* **393**, 643–647 (1998).

Acknowledgements. We thank S. Walker (NOAA/PMEL) for her assistance in development of the MAPR instruments and the Captain, crew and IFRTF team of the RV *Marion Dufresne* who contributed to the success of this project. The FUJI cruise was co-funded by IFRTF (France) and The Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture (Japan). TOBI support was funded through the EC TMR Grant: European Access to Seafloor Survey Systems (EASSS) whilst hydrothermal research was supported by NERC and the SOC (UK) and by the NOAA-Vents program (USA).

Correspondence and requests for materials should be addressed to C.R.G. (e-mail: Chris.R.German@soc.soton.ac.uk).

Electrical conductivity of silicate perovskite at lower-mantle conditions

Tomoo Katsura, Kiminori Sato* & Eiji Ito

Institute for Study of the Earth's Interior, Okayama University, Misasa, Tottori-ken 682-0193, Japan

Geophysical models of the electrical conductivity of the Earth's mantle based on the observed variations of electric and magnetic fields at the surface of the Earth yield estimates of about 1 S m^{-1} for the conductivity of the uppermost lower mantle^{1,2}. But laboratory conductivity measurements on silicate perovskite (thought to be the dominant constituent of the lower mantle) at high pressures have given conflicting estimates of mantle conductivity, ranging from less than 10^{-5} up to 1 S m^{-1} (refs 3–6). Here we present measurements of the electrical conductivity of perovskite in a multi-anvil press at conditions appropriate for the uppermost lower mantle (pressures up to 23 GPa and temperatures up to 2,000 K). We find that the geophysical estimate of lower-mantle electrical conductivity can be well explained by the conductivity of the perovskite component of a low-oxygen-fugacity mantle composed of pyrolite⁷ (the assemblage of mineral phases thought to broadly represent that of the Earth's mantle), assuming a standard geotherm. Our results also indicate that the temperature dependence of perovskite conductivity at lower-mantle temperatures and pressures is significantly larger than shown previously; extrapolations of low-temperature conductivity measurements to the higher temperatures of the lower mantle should therefore be treated with caution.

The starting material, $\text{Mg}_{0.93}\text{Fe}_{0.07}\text{SiO}_3$ enstatite, was synthesized from Mg and Fe metals and tetraethylorthosilicate at 1,450 K for 3 h at an oxygen partial pressure near the iron–wüstite buffer. The broad X-ray diffraction peaks suggest that this enstatite was poorly crystallized and highly reactive. For a pyrolitic mantle⁷, perovskite of this composition is expected at the top of the lower mantle, on the basis of Fe–Mg partitioning experiments between perovskite and magnesiowüstite⁸.

Details of the experimental method have been given elsewhere⁹. The powdered sample was embedded in an Al_2O_3 ring and sandwiched between Fe disks. The oxidation state of the sample was kept at or below the iron–wüstite buffer. The sample is connected in series with a reference resistance of $101 \text{ k}\Omega$. A sinusoidal a.c. voltage (0.01 Hz and $1 V_{\text{peak-to-peak}}$) was applied to this circuit, and the voltages on the sample and the reference were monitored by two digital voltmeters. The impedance of the sample was obtained from the product of the reference resistance and the ratio of the sample to reference voltage.

Only one experiment succeeded at a pressure of 23 GPa. In the first heating cycle of this experiment, the conductivity measure-

ments were made in 100 K steps from 300 to 2,000 K. After the measurement at 2,000 K, the sample was quenched, and measurements were repeated from 300 K to higher temperatures. The measurement in the second cycle was stopped at 900 K; we have to avoid back-transformation of perovskite to ilmenite in the second cycle, and confirm formation of perovskite in the first cycle by means of X-ray diffractometry after measurement.

The experimental results are shown in Fig. 1. In the first heating cycle, the conductivity gradually increased with increasing temperature from 300 to 1,000 K. From 1,100 to 1,300 K, the conductivity increased rapidly, and then dropped off slightly. In the range 1,500–2,000 K, the conductivity rapidly increased again; the logarithm of the conductivity increases linearly with decreasing reciprocal temperature. On quenching from 2,000 to 310 K, the nominal conductivity value dropped by five orders of magnitude. In the second heating cycle, the conductivity increased gradually up to 400 K. However, we suspect that the values over this temperature range are not the sample conductivity, because the nominal sample resistance is comparable to the insulation resistance of the assembly in this conductivity range⁹. Above 500 K, the logarithm of conductivity increased linearly with decreasing reciprocal temperature.

Phases present were not identified *in situ* during the measurement. However, the recovered sample was confirmed to be perovskite by X-ray diffractometry. The change in slope of the electrical conductivity in the first heating cycle can be reasonably attributed to the following series of phase transitions. Up to 1,000 K, the enstatite starting material persisted, but it began to transform to ilmenite at $\sim 1,100 \text{ K}$. The ilmenite then transformed to perovskite at temperatures from 1,300 to 1,500 K. This perovskite, subsequently quenched from 2,000 K, persisted at temperatures of 310–900 K in the second heating cycle.

The recovered sample was also examined by scanning electron microscopy and electron probe microanalysis. These analyses did not show any evidence of reaction of the sample with the Al_2O_3 sample container. The inertness of perovskite with corundum is implied by the very small grain sizes (of the order of $0.1 \mu\text{m}$) of perovskite and corundum synthesized from the homogeneous $\text{MgSiO}_3\text{--Al}_2\text{O}_3$ glasses¹⁰.

As shown in Fig. 1, the logarithm of the conductivity of perovskite in its stability field (1,500–2,000 K in the first cycle) has a steeper slope than that of the quenched 'metastable' perovskite (500–900 K in the second cycle). The electrical conductivity σ can be expressed as follows:

$$\sigma = \sigma_0 \exp(-E/kT) \quad (1)$$

where σ_0 is the conductivity at infinitely high temperature, E is the activation energy, and k is the Boltzmann constant. The activation energies for the stable and quenched perovskite are found to be 0.92 and 0.41 eV, respectively. The activation energy for the quenched perovskite obtained here is in good agreement with that reported by Shankland *et al.*⁶ (0.42 eV at 23 GPa), who also measured the conductivity of quenched perovskite at similar temperatures. However, our results show that conductivity of perovskite obtained at such low temperatures should not be extrapolated to estimate conductivity at high temperatures. The electrical conductivity of perovskite must be measured in its stability field in order to make direct comparisons with the conductivity of the lower mantle.

The reason for the different activation energy in the high- and low-temperature regimes is not clear. Usually, such change is attributed to a change in the conduction mechanism. For example, in many semiconductors, extrinsic conduction mechanisms, such as impurities, have low activation energy and are dominant at lower temperatures, whereas intrinsic mechanisms have high activation energy and become dominant at higher temperatures. One possible explanation for the present case is that the conduction mechanism at temperatures of 400–900 K is electrons hopping from Fe^{2+} to Fe^{3+}

* Present address: National Institute for Research in Inorganic Materials, Tsukuba, Japan.

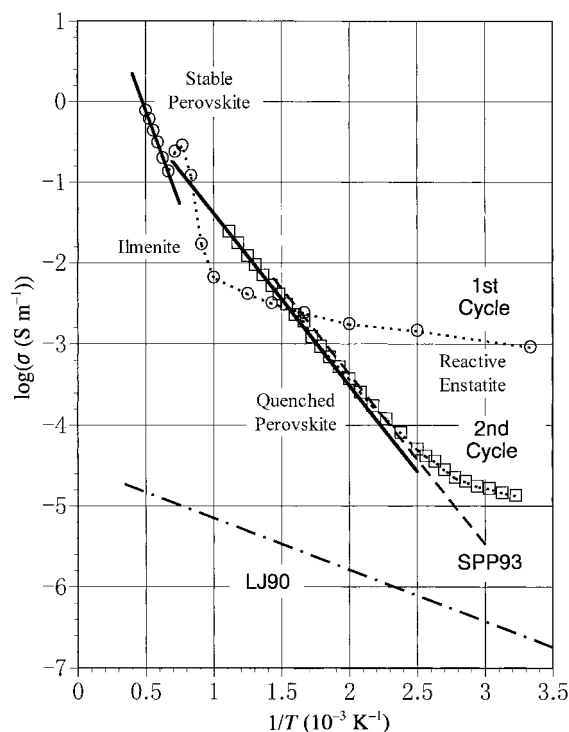


Figure 1 Electrical conductivity σ of stable and quenched perovskite with $(\text{Mg}_{0.93}\text{Fe}_{0.07})\text{SiO}_3$ composition. The circles and squares correspond to conductivity values measured in the first and second heating cycles, respectively (see text). In the first cycle, the sample should be perovskite at temperatures $>1,500$ K; in the second cycle, it should be perovskite over the whole temperature range, although the values were affected by the insufficient electrical insulation of the assembly. The conductivity values obtained by the previous measurements in a diamond anvil cell^{4,6} are also shown for comparison. LJ90, Li and Jeanloz⁴ for $(\text{Mg}_{0.88}\text{Fe}_{0.12})\text{SiO}_3$ perovskite at 57 GPa; SPP93, Shankland *et al.*⁶ for $(\text{Mg}_{0.89}\text{Fe}_{0.11})\text{SiO}_3$ perovskite at 23 GPa.

as suggested by Shankland *et al.*⁶, and that the mechanism at 1,500–2,000 K is ionic.

Another possible explanation for the change in slope is that an increase in carrier density at high temperatures is responsible for the steeper slope, although the conduction mechanism remains the same over both the low- and high-temperature regimes. In the second heating cycle, the density of point defects in perovskite at 2,000 K should remain constant on quenching because of the low temperatures. The electrical conductivity of the quenched perovskite should be proportional to $\exp(-E_a/kT)$, where E_a is the activation energy of electrical conduction, k is the Boltzmann constant, and T is the absolute temperature. Conversely, in the case of stable, high-temperature perovskite, the density of point defects should be proportional to $\exp(-E_f/kT)$, where E_f is the formation energy of a point defect. Therefore, the conductivity of perovskite at high temperatures should be proportional to $\exp(-(E_a + E_f)/kT)$, resulting in an increase in slope by E_f over that of the quenched perovskite at low temperature.

We calculate σ_0 values of 2×10^2 and 4 S m^{-1} for stable and quenched perovskite, respectively. The values of E and σ_0 obtained here are in significant disagreement with those reported by Li and Jeanloz^{3,4}. The sample environment in a multi-anvil press is much better controlled than in a laser-heated diamond anvil cell, and this probably contributed to the differences between the studies. For example, the oxidation state, which is one of the most important factors affecting the electrical conductivity of ferromagnesian silicates¹¹, was constrained by Fe disks in this study. However, in the experiments of Li and Jeanloz, the oxidation state of the sample was not controlled in any way.

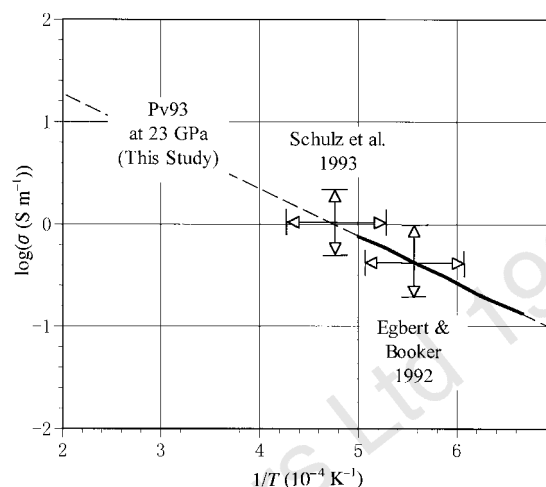


Figure 2 Electrical conductivity σ of $(\text{Mg}_{0.93}\text{Fe}_{0.07})\text{SiO}_3$ perovskite in the upper part of the lower mantle. The solid line denotes the electrical conductivity of $(\text{Mg}_{0.93}\text{Fe}_{0.07})\text{SiO}_3$ perovskite measured in this study. The dashed line denotes extrapolation of the present measurement results. The vertical arrows show the ranges of electrical conductivity of the lower mantle at 700 km depth in the geophysical models¹²; the horizontal arrows indicate the ranges of temperature at the top of the lower mantle estimated from the present measurement and the geophysical models¹².

The electrical conductivity of perovskite obtained in this study is compared with two recent geomagnetic models for the Earth's mantle^{1,2} in Fig. 2. Both of the geomagnetic models are consistent with our measured conductivity values of $(\text{Mg}_{0.93}\text{Fe}_{0.07})\text{SiO}_3$ perovskite under an oxidation state close to the iron–wüstite buffer for the temperature range 1,800–2,100 K. Thus, we conclude that the electrical conductivity at the uppermost part of the lower mantle can be reasonably explained by a reduced pyrolytic model mantle composition at temperatures consistent with standard model geotherms^{8,12}.

In this study, the pyrolytic model mantle is simplified as $(\text{Mg}_{0.93}\text{Fe}_{0.07})\text{SiO}_3$ perovskite. However, there are two factors that may affect lower-mantle conductivity. One is the presence of an additional lower-mantle mineral, magnesiowüstite. However, it occupies only 17 vol.% in the pyrolytic model mantle; we cannot expect that magnesiowüstite forms a network structure, as required to provide an effective conduction path, in the lower mantle. This is supported by the textural observations for the perovskite–magnesiowüstite assemblage reported by Martinez *et al.*¹⁴. Therefore, we consider that the effect of magnesiowüstite on the lower-mantle conductivity should be small.

A second factor that may affect lower-mantle conductivity is the possible presence of ferric iron in perovskite in this region. Although perovskite at the top of the lower mantle will not contain aluminium, the concentration of Al_2O_3 gradually increases with increasing depth up to 720 km (ref. 15). McCammon¹⁶ showed that Al-bearing perovskite should contain some ferric iron even under reduced conditions. This implies that lower-mantle perovskite could have higher conductivity than measured in this study. However, in comparison with the uppermost lower mantle, the remaining lower mantle has higher conductivity^{1,2}. We suggest that an increase in Fe^{3+} in perovskite in the lower mantle can produce good agreement of the mineral physical data with the geomagnetic observations. □

Received 27 October 1997; accepted 14 July 1998.

1. Egbert, G. D. & Booker, J. R. Very long period magnetotellurics at Tucson observatory: implication for mantle conductivity. *J. Geophys. Res.* **97**, 15099–15112 (1992).
2. Schulz, A., Kurtz, R. D., Chave, A. D. & Jones, A. G. Conductivity discontinuity in the upper mantle beneath a stable craton. *Geophys. Res. Lett.* **20**, 2941–2944 (1993).

3. Li, X.-Y. & Jeanloz, R. Electrical conductivity of (Mg,Fe)SiO₃ perovskite and perovskite-dominated assemblage at lower mantle conditions. *Geophys. Res. Lett.* **14**, 1075–1078 (1987).
4. Li, X.-Y. & Jeanloz, R. Laboratory studies of the electrical conductivity of silicate perovskites at high pressures and high temperatures. *J. Geophys. Res.* **95**, 2067–2078 (1990).
5. Peyronneau, J. & Poirier, J. P. Electrical conductivity of Earth's lower mantle. *Nature* **342**, 537–539 (1989).
6. Shankland, T. J., Peyronneau, J. & Poirier, J. P. Electrical conductivity of Earth's lower mantle. *Nature* **366**, 453–455 (1993).
7. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, New York, 1979).
8. Katsura, T. & Ito, E. Determination of Fe-Mg partitioning between perovskite and magnesiowüstite. *Geophys. Res. Lett.* **23**, 2005–2008 (1996).
9. Katsura, T., Sato, K. & Ito, E. Electrical conductivity measurement of minerals at high pressures and high temperatures. *Rev. High Pressure Sci. Technol.* **7**, 18–21 (1998).
10. Ito, E., Kubo, A., Katsura, T., Akaogi, M. & Fujita, T. High-pressure transformation of pyrope (Mg₃Al₂Si₂O₁₂) in a sintered diamond cubic anvil assembly. *Geophys. Res. Lett.* **25**, 821–824 (1998).
11. Wood, B. J. & Nell, J. High-temperature electrical conductivity of the lower-mantle phase (Mg,Fe)O. *Nature* **351**, 309–311 (1991).
12. Ito, E. & Katsura, T. A temperature profile in the mantle transition zone. *Geophys. Res. Lett.* **16**, 425–428 (1989).
13. Dobson, D. P., Richmond, N. C. & Brodholt, J. P. A high-temperature electrical conduction mechanism in the lower mantle phase (Mg,Fe)_{1-x}O. *Science* **273**, 1779–1781 (1997).
14. Martínez, I., Wang, Y., Guyot, F., Liebermann, R. C. & Doukhan, J.-C. Microstructures and iron partitioning in (Mg,Fe)SiO₃ perovskite-(Mg,Fe)O magnesiowüstite assemblages: an analytical transmission electron microscopy study. *J. Geophys. Res.* **102**, 5256–5280 (1997).
15. Irfune, T. Absence of an aluminous phase in the upper part of the Earth's lower mantle. *Nature* **370**, 131–133 (1994).
16. McCammon, C. Perovskite as a possible sink for ferric iron in the lower mantle. *Nature* **387**, 694–696 (1997).

Acknowledgements. We thank M. J. Walter and J. Brodholt for comments.

Correspondence and requests for materials should be addressed to T.K. (e-mail:tkatsura@misasa.okayama-u.ac.jp).

Productivity controls food-chain properties in microbial communities

Christina M. K. Kaunzinger & Peter J. Morin

Department of Ecology, Evolution & Natural Resources, Rutgers University, New Brunswick, New Jersey 08903-0231, USA

Food webs provide a framework for integrating population dynamics, community structure and ecosystem processes, but the causes of many food-web patterns remain controversial¹. An unresolved issue concerns the factors that limit food-chain length in natural ecosystems, although theory predicts that length could be limited by inefficient energy transfer between trophic levels² or by the instability of longer food chains³. Experiments provide qualitative support for the instability of longer food chains⁴, but, according to theory, this relationship may depend on the assumptions used to model food chains⁵. Here we show that experimental manipulations of productivity determine the length of microbial food chains in laboratory microcosms. Food-chain length is also predicted to determine how productivity influences population densities within trophic levels. If trophic levels consist entirely of edible prey, and if consumers feed only on the next-lower trophic level⁶, population densities should alternate between increasing and constant values as food-chain length increases (Fig. 1). We find that food-chain length determines population-level responses to productivity, which is consistent with predictions from models. These results indicate that the impacts of human alterations of productivity will depend crucially on the structure of the affected food chains.

Unambiguous tests of food-web theory are difficult to do in natural systems. The problems of measuring the length of food chains accurately, and of assigning species to single trophic levels, limit the interpretation of field studies of the impact of productivity on natural food chains^{7,8}. Long response times of long-lived organisms can also delay responses to manipulations of productivity. Laboratory studies of microbial food chains can overcome the obstacles presented by field studies of long-lived, omnivorous

organisms. Short generation times of microbes (3–48 h) ensure rapid population responses to altered productivity, and food chains can be constructed so that organisms occupy single trophic levels⁹.

We did two experiments to study how the properties of linear food chains changed along gradients of increasing productivity. Food chains consisted of aquatic bacteria and protists. We manipulated productivity by altering the nutrient concentration of a standard growth medium. The first experiment described how increased productivity affected population densities in food chains with one or two trophic levels. The bacterium *Serratia marcescens* occupied the first or basal trophic level, and the bacterivorous ciliate *Colpidium striatum* occupied the second trophic level. Twelve different nutrient concentrations established a productivity gradient spanning nearly two orders of magnitude (Fig. 2). A second experiment used food chains with initially one, two or three trophic levels to test whether longer food chains would persist at higher levels of productivity only. *Didinium nasutum*, a predatory ciliate that consumes *Colpidium*, occupied the third trophic level. The first level of the two- and three-level chains contained either one (in the two-level chain) or several (in the three-level chain) bacterial taxa in addition to *Serratia*. These bacteria entered the focal chains as unavoidable contaminants with *Colpidium* and *Didinium*, but they did not change key features of food-web architecture, or the

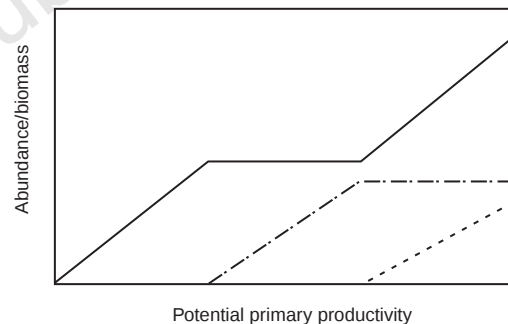


Figure 1 The pattern of abundance within trophic levels expected in response to increases in productivity in food-chain models. Solid line, trophic level 1; dotted and dashed line, trophic level 2; dotted line, trophic level 3. Modified from ref. 20.

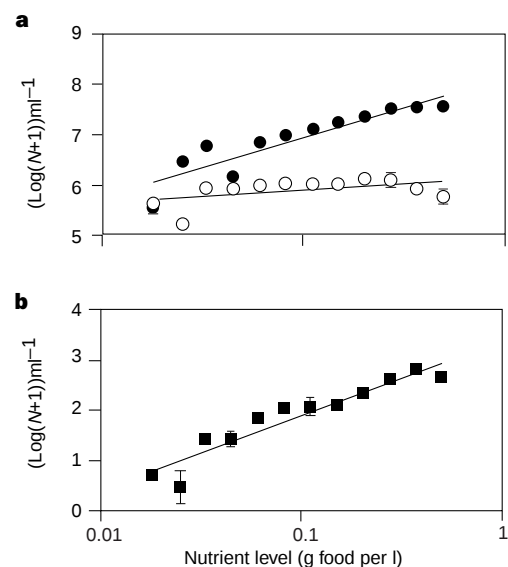


Figure 2 Effects of productivity on population abundances in trophic levels 1 and 2 of experiment 1. **a**, Response of *Serratia* abundance to a productivity gradient when cultured alone (filled circles) or with *Colpidium* (open circles). **b**, Response of *Colpidium* abundance to a productivity gradient. Lines were fitted by regression; error bars indicate s.e.m. Productivity scales with g food per litre, which indicates the concentration of protozoan pellet in the nutrient medium.

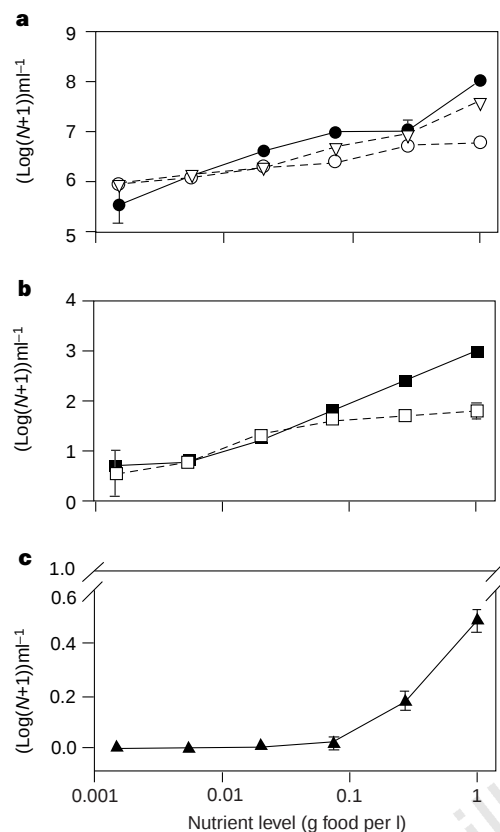


Figure 3 Effects of productivity on population abundances in trophic levels 1–3 of experiment 2. **a**, Response of bacterial abundance to a productivity gradient when cultured alone (filled circles), with *Colpidium* (open circles), or with *Colpidium* and *Didinium* (triangles). **b**, Response of *Colpidium* abundance to a productivity gradient when cultured alone (filled squares) or with *Didinium* (open squares). **c**, Response of *Didinium* abundance to a productivity gradient. Error bars indicate s.e.m. Productivity scales with g food per litre.

responses of basal trophic levels to productivity or consumers (Figs 2a and 3a). Six different nutrient concentrations varied productivity over nearly three orders of magnitude (Fig. 3).

The longest (three-level) food chains persisted only at the three highest values of productivity (Fig. 3c and Table 1), with *Didinium* populations attaining maximal size at the highest productivity levels. Extinctions of *Didinium* populations at lower productivity levels show the positive effects of productivity on food-chain length. Consequently, direct or indirect effects of *Didinium* on lower trophic levels appeared only at higher productivity levels (Fig. 3a, b, and Table 1). Two-level food chains, resulting either from the loss of *Didinium* or by design, persisted across the entire productivity gradient in both experiments.

Food-chain length determined whether population densities

within trophic levels increased substantially as productivity increased. Bacterial densities (the first trophic level) increased with productivity in all food chains without higher trophic levels (Figs 2a, 3a and Table 1). Grazing by *Colpidium* greatly reduced the slope of the relation between productivity and bacterial abundance (Figs 2a, 3a and Table 1) in two-level food chains. *Colpidium* abundance increased markedly with increased productivity (Figs 2b, 3b and Table 1) in the absence of higher trophic levels, whether the third trophic level was absent by design or as a consequence of low productivity. Consumers in the third trophic level (*Didinium*) decreased the slope of the relation between productivity and *Colpidium* abundance (Fig. 3b and Table 1), and increased the slope of the relation between productivity and bacterial abundance (Fig. 3a and Table 1), over the portion of the gradient that supported three-level food chains. Observed changes in the slope of productivity–abundance relations within trophic levels that coincide with changes in food-chain length are consistent with food-chain theory⁶.

To our knowledge, this is the first experimental study to link productivity-induced changes in food-chain length with cascading changes in population density at several trophic levels. Previous work focused either on the effects of productivity on food-chain length^{10,11}, or on the effects of food-chain length on trophic cascades^{12–15}. Comparisons of lakes¹¹ and experiments with tree-hole communities¹⁰ indicate that food-chain length increases with productivity. Studies of lakes^{12,13}, streams¹⁴ and chemostats¹⁵ also show that responses of population density to productivity depend on food-chain length. Both our study and others found a slight but significant increase in bacterial abundance in the two-level food chains, which is not predicted by classical models. This discrepancy could result from several biologically realistic complications, including spatial heterogeneity¹⁵, predator interference¹⁶, or evolutionary or behavioural responses of the prey to predators¹⁷. Because we designed our experiments to minimize complications caused by prey heterogeneity^{18,19} and omnivory^{20,21}, neither of these factors should generate positive correlations between resources and consumers. Ratio-dependent models could produce the positive relations observed between consumers and resources²², but such interactions seem unlikely in most systems^{15,23}. Some experimental studies also describe responses of lower trophic levels that depart from patterns suggested by food-chain models²⁴.

Laboratory microcosms provide an essential bridge between mathematical models^{6,25}, which are both abstracted and simplified, and the full complexity of nature. Our food chains corresponded to the linear basal portions of more reticulate, natural food webs. The qualitative agreement between model predictions and food-chain properties indicates that departures of field studies from model predictions reflect important differences in the architecture of natural and model food chains. Our approach can test this hypothesis by building more complex food chains and observing their responses to differences in productivity. Future studies should vary trophic complexity and omnivory to determine how these aspects of natural systems alter responses to changes in productivity. □

Table 1 Relation between population abundance and experimentally imposed productivity gradients

Species	FCL	Experiment 1	Experiment 2	
		All nutrient levels Slope (s.e.m.)	Nutrient levels 1–3 Slope (s.e.m.)	Nutrient levels 4–6 Slope (s.e.m.)
Bacteria	1	1.17 (0.12)***	0.95 (0.23)*	0.92 (0.27)*
Bacteria	2	0.26 (0.10)*	0.32 (0.03)**	0.34 (0.07)**
Bacteria	3		0.32 (0.06)**	0.81 (0.10)**
<i>Colpidium</i>	2	1.50 (0.12)***	0.46 (0.13)*	1.03 (0.06)***
<i>Colpidium</i>	3		0.67 (0.30)	0.17 (0.11)
<i>Didinium</i>	3		0.01 (0.002)*	0.41 (0.05)**

Slopes for linear regressions of the relationship between population abundance in food chains of different length and experimentally imposed productivity gradients are shown. Separate slope estimates for nutrient levels 1–3 and 4–6 in experiment 2 correspond to portions of the gradient that support two and three trophic levels, respectively. FCL, initial food-chain length. Levels of significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$.

Methods

We used standard culture methods for protists to establish simple microbial food chains in microcosms⁴. Treatment of *Colpidium* with a penicillin–streptomycin–neomycin solution eliminated the bacterial assemblage of the stock culture. In experiment 1 we used 12 nutrient levels, ranging from 0.018 to 0.5 g protozoan pellet per litre of well water, evenly spaced on a log scale, with two replicates of each treatment. The ‘*Serratia* alone’ combination ran for 33 days in 15 ml filtered protist pellet medium in 20-ml screw-cap test tubes. *Serratia* was plated every four days during the initial part of the experiment (days 5–17) and every eight days thereafter. Removal and replacement of 4% of the medium every four days renewed nutrients. Counts of *Colpidium* at four-day intervals used a 0.25-ml (± 0.06 ml) subsample from the removed medium. The ‘*Serratia* + *Colpidium*’ combination ran for 25 days, or ~ 100 *Colpidium* generations.

In experiment 2 we used six nutrient levels ranging from 0.0015 to 1 g protozoan pellet per litre of well water, evenly spaced on a log scale, with two replicates of each treatment. Treatment of *Colpidium* with a penicillin–streptomycin–neomycin solution eliminated all but one bacterial contaminant. Antibiotics could not be used to eliminate bacteria from *Didinium* stock cultures because of toxicity. Food-chain combinations ran for 40 days in 100 ml unfiltered protist pellet medium in 250-ml screw-cap Erlenmeyer flasks. Bacteria were plated three times during the experiment for the two- and three-level combinations and four times for the one-level combination. Removal and replacement of 5% of the medium every four days renewed nutrients. Counts of protists at four-day intervals used a subsample of 0.1–3.5 ml (*Colpidium*, subsample size depended upon protist density) or of 0.2–4.6 ml (*Didinium*) from the medium withdrawn for nutrient replacement.

Received 2 March; accepted 3 July 1998.

- Pimm, S. L., Lawton, J. H. & Cohen, J. E. Food web patterns and their consequences. *Nature* **350**, 669–674 (1991).
- Hutchinson, G. E. Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
- Pimm, S. L. & Lawton, J. H. Number of trophic levels in ecological communities. *Nature* **268**, 329–331 (1977).
- Lawler, S. P. & Morin, P. J. Food web architecture and population dynamics in laboratory microcosms of protists. *Am. Nat.* **141**, 675–686 (1993).
- Sterner, R. W., Bajpai, A. & Adams, T. The enigma of food chain length: absence of theoretical evidence for dynamic constraints. *Ecology* **78**, 2258–2262 (1997).
- Oksanen, L., Fretwell, S. D., Arruda, J. & Niemela, P. Exploitation ecosystems in gradients of primary productivity. *Am. Nat.* **118**, 240–261 (1981).
- Power, M. E. Top-down and bottom-up forces in food webs: do plants have primacy? *Ecology* **73**, 733–746 (1992).
- Polis, G. A. & Strong, D. R. Food web complexity and community dynamics. *Am. Nat.* **147**, 813–846 (1996).
- Morin, P. J. & Lawler, S. P. Food web architecture and population dynamics: theory and empirical evidence. *Annu. Rev. Ecol. Syst.* **26**, 505–529 (1995).
- Jenkins, B., Kitching, R. L. & Pimm, S. L. Productivity, disturbance and food web structure at a local spatial scale in experimental container habitats. *Oikos* **65**, 249–255 (1992).
- Persson, L., Diehl, S., Johansson, L., Andersson, G. & Hamrin, S. F. Trophic interactions in temperate lake ecosystems: a test of food chain theory. *Am. Nat.* **140**, 59–84 (1992).
- Hansson, L.-A. The role of food chain composition and nutrient availability in shaping algal biomass development. *Ecology* **73**, 241–247 (1992).
- Mazumder, A. Patterns of algal biomass in dominant odd- vs. even-link lake ecosystems. *Ecology* **75**, 1141–1149 (1994).
- Wootton, J. T. & Power, M. E. Productivity, consumers, and the structure of a river food chain. *Proc. Natl Acad. Sci. USA* **90**, 1384–1387 (1993).
- Bohannan, B. J. M. & Lenski, R. E. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. *Ecology* **78**, 2303–2315 (1997).
- McCann, K. S., Hastings, A. & Strong, D. R. Trophic cascades and trophic trickles in pelagic food webs. *Proc. R. Soc. Lond. B* **265**, 205–209 (1998).
- Moen, J. & Oksanen, L. Ecosystem trends. *Nature* **353**, 510 (1991).
- Leibold, M. A. Resource edibility and the effects of predators and productivity on the outcome of trophic interactions. *Am. Nat.* **134**, 922–949 (1989).
- Abrams, P. A. Effect of increased productivity on the abundance of trophic levels. *Am. Nat.* **141**, 351–371 (1993).
- Mittelbach, G. G., Osenberg, C. W. & Leibold, M. A. In *Size-Structured Populations: Ecology and Evolution* (eds Ebenman, B. & Persson, L.) 219–235 (Springer, Berlin, 1988).
- Polis, G. A. Complex trophic interactions in deserts: an empirical critique of food web theory. *Am. Nat.* **138**, 123–155 (1991).
- Arditi, R. & Ginzburg, L. R. Coupling in predator–prey dynamics: ratio-dependence. *J. Theor. Biol.* **139**, 311–326 (1989).
- Abrams, P. A. The fallacies of “ratio-dependent” predation. *Ecology* **75**, 1842–1850 (1994).
- Brett, M. T. & Goldman, C. R. Consumer versus resource control in freshwater pelagic food webs. *Science* **275**, 384–386 (1997).
- Leibold, M. A. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. *Am. Nat.* **147**, 784–812 (1996).

Acknowledgements. We thank M. Cole, J. Fox, P. Harris, M. Holyoak, L. Ginzburg, S. Lawler, M. Leibold, J. McGrady-Steed, M. Power, V. Schmalhofer and D. Strong for comments on the manuscript. This work was supported by the NSF and the Leatham and Busch Funds of Rutgers University.

Correspondence and requests for materials should be addressed to C.M.K.K. (e-mail: cmkk@eden.rutgers.edu).

Terrain influences the accurate judgement of distance

Michael J. Sinai*, Teng Leng Ooi† & Zijiang J. He*

* Department of Psychology, University of Louisville, Louisville, Kentucky 40292, USA

† Department of Biomedical Sciences, Southern College of Optometry, Memphis, Tennessee 38104, USA

Mathematically, three-dimensional space can be represented differently by the cartesian, polar, and other coordinate systems. However, in physical sciences, the choice of representation system is restricted by the need to simplify a machine’s computation while enhancing its efficiency¹. Does the brain, for the same reasons, ‘select’ the most cost-efficient way to represent the three-dimensional location of objects? As we frequently interact with objects on the common ground surface, it might be beneficial for the visual system to code an object’s location using a ground-surface-based reference frame². More precisely, the brain could use a quasi-two-dimensional coordinate system (x_s, y_s) with respect to the ground surface (s), rather than a strictly three-dimensional coordinate system (x, y, z), thus reducing coding redundancy and simplifying computations^{2–5}. Here we provide support for this view by studying human psychophysical performance in perceiving absolute distance and in visually directed action tasks^{6–11}. For example, when an object was seen on a continuous, homogeneous texture ground surface, the observer judged the distance to the object accurately. However, when similar surface information was unavailable, for example, when the object was seen across a gap in the ground, or across distinct texture regions, distance judgement was impaired.

The idea of representing space using the ground surface as the reference frame is fundamental to the ‘ecological approach to space perception’ pioneered by J. J. Gibson². A vital prediction of Gibson’s ‘ground theory’² is that, when the common ground surface is disrupted, the visual system is unable to establish a reliable reference frame and consequently fails to obtain correct absolute distance. To test this theory, we placed a target at the other side of a gap in the ground surface (0.5 m deep and 1.3 m wide) from a naive observer, whose task was to judge the absolute distance of the target from himself (Fig. 1a). Then he was blindfolded or asked to shut his eyes, turned 90° away, and instructed to walk a distance equivalent to the remembered absolute distance of the target. Ten observers participated and the average distance walked was 4.60 ± 0.12 m (Fig. 1a), which was significantly farther than the target’s physical distance, 3.66 m ($t(9) = 8.2, P < 0.001$). As a control, we tested five observers over the same physical distance on a continuous surface (with no gap) and found that the average distance walked was 3.69 ± 0.12 m (Fig. 1a), in close agreement with the target’s physical distance from the observer ($t(4) = 0.3, P > 0.05$).

We then determined whether the common ground surface, in addition to affecting visually directed action, also influences conscious distance perception. We asked five naive observers to judge the absolute distance of a target under similar viewing conditions, and then to perceptually set the distance of a matching target to be at an equal distance. We found that the average matched distance was 4.24 ± 0.6 m when the target was seen across a gap (Fig. 1a) indicating that the observers overestimated the absolute distance ($t(4) = 10.2, P < 0.001$). Conversely, the average matched distance for the continuous surface condition was 3.54 ± 0.07 m (Fig. 1a), which was very close to the physical distance, 3.66 m ($t(4) = -1.44, P > 0.05$). The resemblance between the perceptual matching results and the blindfolded-walking results indicates that the inaccuracy in absolute distance judgement on a discontinuous

surface occurs for both tasks¹². This implies that the internal representation of space, which is based on the common ground surface, is used for both space perception and visually directed performance^{11,13}.

To test the ground theory further, we repeated both the blindfolded walking and the perceptual matching tasks under more stringent conditions. The gap size was now widened to 4.1 m and deepened to 2.0 m (Fig. 1b). Consistent with earlier results, the observers ($n = 8$) also overestimated the distance to target under the gap condition ($t(7) = 5.03$, $P < 0.01$ (walking); $t(7) = 12.4$, $P < 0.01$ (perceptual matching)), and performed quite accurately in the continuous surface condition ($t(7) = -0.32$, $P > 0.05$ (walking); $t(7) = 0.24$, $P > 0.05$ (perceptual matching)).

In all, our results provide evidence for the role of the common ground surface in accurate absolute distance judgement. But what depth cues on the common surface are used to establish its reference frame? As blindfolded walking can be accurate up to 12 m on a continuous ground surface⁷⁻¹⁰, we proposed that the texture gradient of size could be among the likely cues for providing quantitative distance information¹⁴. To test this possibility, we began by consulting a distance/texture model⁵ which states that,

with some approximation, the perceived absolute distance, Z , equals $(H \times G)/3$, where H is the observer's eye height relative to the ground surface, and G is the local texture gradient of size on the ground at the target's location. If the brain uses this strategy, the perceived absolute distance on the common ground surface should depend on both H and G .

To determine the impact of H on perceived absolute distance, our naive observers stood on an elevated ground surface and estimated horizontal distance of a target placed on a lower ground surface (dashed arrow, Fig. 2a). Eight observers performed the blindfolded-walking task while five observers performed the perceptual distance matching task. Their results show overestimation (Fig. 2a, diamond and circle symbols, respectively).

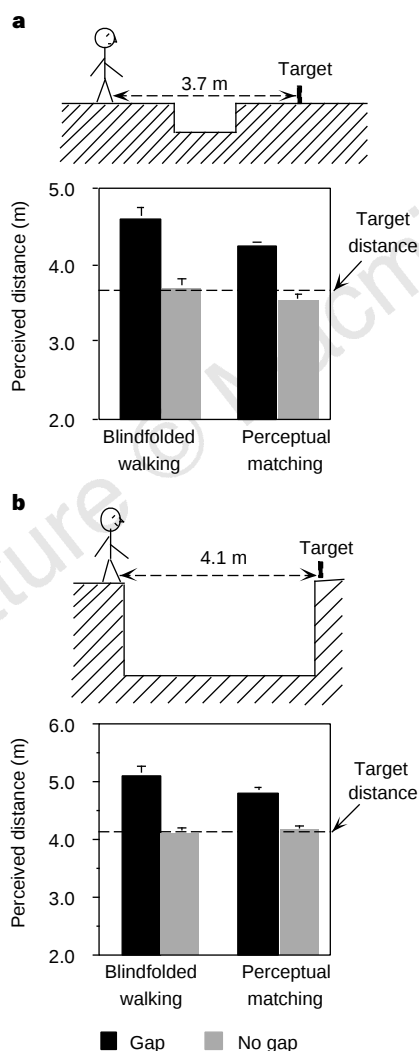


Figure 1 The effect of surface discontinuity (gap) on perceived absolute distance. **a**, A gap in the ground 0.5 m deep and 1.3 m wide separates the target and observer. **b**, The gap was increased to 2.0 m deep and 4.1 m wide. The graph below each illustration plots the average results from the blindfolded-walking and perceptual matching tasks for that condition. The black bars represent the data from the gap condition, and the grey bars data from the control condition, in which the ground surface was continuous (no gap).

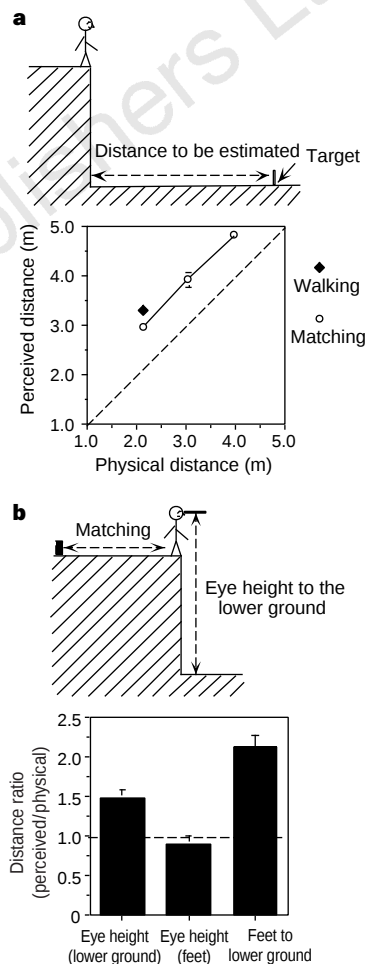


Figure 2 Eye height and surface elevation. **a**, The effect of elevated ground surface on distance perception. The observer stood on the elevated ground surface (2.0 m from the lower ground surface) and estimated the distance of the target, placed on the lower ground surface, from him (arrow). The graph below plots the perceived distance as a function of the target's physical distance from the observers for the blindfolded walking (diamond) and perceptual matching (circle) tasks. Distance overestimation is evidenced by the fact that the data points are located above the dashed line, which demarcates equal physical and perceived distances. The standard error of each data point is represented by the vertical error bar. (Error bars are not shown when they are smaller than the symbol.) **b**, Perception of eye height. The illustration shows the perceptual matching task for measuring the eye height with respect to the lower ground surface. The average result (ratio of matched distance to physical height) is depicted by the left bar in the graph. The middle bar represents the ratio of the eye height with respect to the feet to the physical distance (illustration not shown), and the right bar represents the ratio of the perceived elevated height (feet to the lower ground) to the physical distance (illustration not shown).

A likely source of the overestimation could be related to an exaggeration of the observer's perception of the eye height with respect to the lower ground level, H . To study this possibility, we asked five naive observers to stand on the higher ground surface and to perceptually set the distance of a matching target to equal their perceived eye height to the lower ground surface (Fig. 2b). As predicted, the observers overestimated their eye heights with respect to the lower ground surface (left bar, Fig. 2b, $t(4) = 3.72$, $P < 0.05$). Then we used the same procedure to measure the observers' perception of their eye height with respect to their feet, and found that their estimations were reasonably accurate (middle bar, Fig. 2b, $t(4) = -0.96$, $P > 0.05$). We also measured the observers' perception of the distance between their feet level (on the elevated ground) and the lower ground surface (2.0 m), and found that they overestimated the distance (right bar, Fig. 2b, $t(4) = 7.5$, $P < 0.001$).

These results are consistent with our assumption that the distance overestimation seen in Fig. 2a is due to the observers' overestimation of their eye height with respect to the lower ground surface (H). These observations add support to the idea that the observer's eye height provides information necessary for distance judgement^{5,15,16}. It is likely that the adult observer's eye height with respect to the ground, being a constant most of the time, leads the visual system to internalize it as implicit knowledge (a yardstick).

Finally, we studied the influence of texture gradient information on absolute distance judgement. We used a viewing condition in which the ground surface between the observer and target had two distinct texture regions, namely, concrete and a grass field (Fig. 3). The observer stood on the concrete field and pre-viewed the target on the grass field, and then performed the blindfolded-walking task. We measured four different absolute distances. For each absolute distance, the width of the grass field was kept constant at 3.05 m while the width of the concrete field was varied (0.61–4.56 m). The observers ($n = 10$) tended to underestimate the absolute distance as the distance between the observer and target increased (white

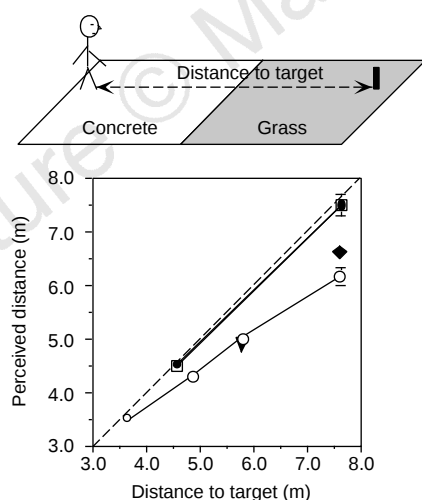


Figure 3 The effect of texture discontinuation on distance perception. White circles represent the distances walked when the observer stood on the concrete field and viewed the target placed on the grass field, as shown in the illustration. The width of the grass field was constant while the width of the concrete field was variable (0.61–4.57 m). The diamond depicts the average result for a perceptual matching task under this viewing condition for a target distance of 7.62 m. The triangle represents the distance walked when the viewing condition was changed so that the observer stood on the grass and viewed the target on a concrete field (viewing distance 5.79 m). In all, the observers showed underestimation of distance (their data being below the dashed line which demarcates equal physical and perceived distances). The filled circles and open squares represent the blindfolded-walking data obtained from the control conditions, under which the common ground surface consisted of homogeneous concrete or grass field, respectively.

circles, Fig. 3). In separate control experiments, we tested observers ($n = 5$) on a homogeneous (grass only or concrete only) texture surface, and found that their average blindfolded walking performance was quite accurate (squares and black circles, Fig. 3). Indeed, the results of these control experiments are consistent with the ground theory.

Then to confirm that the underestimation was not due to the viewing direction being from the concrete field, we repeated the experiment with the observers ($n = 5$) standing on the grass field and viewing the target placed 5.79 m away on the concrete field. Here, too, the observers underestimated the absolute distance by almost the same magnitude (triangle, Fig. 3). This indicates that the underestimation errors are not due to the concrete texture region having a different grain size from the grass texture region. Lastly, to determine whether this inaccuracy in absolute distance judgement also occurs in perception, observers ($n = 5$) performed the perceptual distance matching task at an absolute distance of 7.62 m, while viewing from the concrete to the grass field. On average, the observers underestimated the absolute distance (diamond, Fig. 3). Overall, these results indicate that texture discontinuity causes errors in absolute distance judgement, both for perceptual and for visually directed tasks.

In summary, we have found that texture gradient of size, in addition to eye height, can influence judgement of absolute distance. This indicates that texture gradient on the ground surface acts as a depth cue for the visual system to establish a reference frame. However, whether the absolute distance, Z , is determined by the very specific interaction between eye height and texture gradient of size, $(H \times G)/3$, will require more quantitative studies. No doubt, there are other likely depth cues on the ground (for example, the angular declination below the horizon⁵) that can be used by the visual system for absolute distance computation^{2,5,14,17}. It would be interesting to learn how the various depth cues interact when confronted with diverse terrain conditions, such as those in Figs 1 and 3, to cause an observer systematically to overestimate and underestimate absolute distance, respectively.

More generally, our results support the proposal by J. J. Gibson that the common ground surface is used as a reference frame for coding the location of an object². This has significant biological implications because, as noted earlier, there are many ways in which spatial information can be encoded. The question is, which coding mechanism is the most cost-effective? In other words, what type of neural computation can best reduce coding redundancy and enhance overall efficiency^{2–5}? If the purpose is, ultimately, to ensure the survival of the animal, we are compelled to argue that for terrestrial animals, such as humans, the use of the animal's natural niche, the ground surface, as a reference frame for encoding spatial information is a good start^{2,18}. □

Methods

Observers. Twenty-two naive observers (sixteen males and six females) with self-reported normal vision gave informed consent and participated in the experiments on different days.

Viewing environment. All viewing conditions in daylight (parameters described in text) were carefully selected from the natural landscape within the University campus.

The blindfolded walking task. Because of our particular viewing conditions, we could not use the typical blindfolded walking paradigm used by others^{6–10}. In a typical walking paradigm, the blindfolded observers would walk directly in the direction of the target, whereas in our modified paradigm the observers were asked to turn 90° away before walking the remembered distance. To establish the validity of our modified paradigm, we first tested ten naive observers (who were either blindfolded or instructed to shut their eyes) in pilot experiments, to compare their performances in both paradigms. For the three distances tested, 6.04 m, 8.08 m and 10.58 m, the corresponding average distances walked were 5.90 ± 0.07 m, 7.85 ± 0.17 m and 9.83 ± 0.16 m, respectively, for the typical blindfolded-walking task, and 5.82 ± 0.10 m,

7.92 ± 0.12 m and 9.68 ± 0.19 m, respectively, for the modified blindfolded-walking tasks. The similarity between the results obtained from both types of walking task indicates that our modified walking task can also be used to accurately reflect the observer's distance judgement.

The perceptual matching task. The target viewing conditions were similar to those used for the walking task. To obtain the observer's perception of distance, the matching target was placed 90° from the observer. The observer's task was to view the test target, then turn toward the matching target and instruct the experimenter to adjust the location of the matching target until it appeared to be at the same distance from him as the test target.

Each observer underwent a practice session before commencing the experiments. During the proper experiments, the observers were tested under the same condition two to three times, depending on the particular task. When more than one target distance was tested in an experiment, the order of testing was counterbalanced across observers.

Received 11 May; accepted 16 July 1998.

- Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication* (Univ. Illinois Press, Urbana, 1949).
- Gibson, J. J. *The Perception of the Visual World* (Houghton Mifflin, Boston, 1950).
- Barlow, H. B. in *Sensory Communication* (ed. Rosenblith, W.) 217–235 (MIT Press, Cambridge, Massachusetts, 1961).
- Attneave, F. Informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
- Sedgwick, H. A. in *Human and Machine Vision* (eds Rosenthal, A. & Beck, J.) 425–458 (Academic, New York, 1983).
- Thomson, J. A. Is continuous visual monitoring necessary in visually guided locomotion? *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 427–443 (1983).
- Elliot, D. Continuous visual information may be important after all: a failure to replicate Thomson (1983). *J. Exp. Psychol. Hum. Percept. Perform.* **12**, 388–391 (1987).
- Steenhuis, R. E. & Goodale, M. A. The effects of time and distance on accuracy of target-directed locomotion: does an accurate short-term memory for spatial location exist? *J. Motor Behav.* **20**, 399–415 (1988).
- Rieser, J. J., Ashmead, D. H., Talor, C. R. & Youngquist, G. A. Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception* **19**, 675–689 (1990).
- Loomis, J., DaSilva, J., Fujita, N. & Fukushima, S. Visual space perception and visually directed action. *J. Exp. Psychol.* **18**, 906–921 (1992).
- Loomis, J., DaSilva, J., Philbeck, J. W. & Fukushima, S. Visual perception of location and distance. *Curr. Dir. Psychol. Sci.* **5**, 72–77 (1996).
- Jiang, Y. & Mark, L. S. The effect of gap depth on the perception of whether a gap is crossable. *Percept. Psychophys.* **56**, 691–700 (1994).
- Philbeck, J. W. & Loomis, J. M. Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *J. Exp. Psychol. Hum. Percept. Perform.* **23**, 72–85 (1997).
- Cutting, J. E. & Vishton, P. M. in *Handbook of Perception and Cognition: Perception of Space and Motion* Vol. 6 (eds Epstein, W. & Rogers, S.) 69–117 (Academic, San Diego, 1995).
- Mark, L. S. Eye-height scaled information about affordances: a study of sitting and stair climbing. *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 360–370 (1987).
- Warren, W. H. & Whang, S. Visual guidance of walking through apertures: body-scaled information for affordances. *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 371–383 (1987).
- Sedgwick, H. A. Combining multiple forms of visual information to specify contact relations in spatial layout. *SPIE* **1198**, 447–458 (1989).
- He, J. Z. & Nakayama, K. Apparent motion determined by surface layout not by disparity or by 3-dimensional distance. *Nature* **367**, 173–175 (1994).

Acknowledgements. This research was supported in part by a Sloan Research Fellowship from the Alfred P. Sloan Foundation (to Z.J.H.) and by a Faculty Research Grant from SCO (to T.L.O.).

Correspondence and requests for materials should be addressed to Z.J.H. (e-mail: z0he0002@ulkyvm.louisville.edu).

Shape selectivity in primate lateral intraparietal cortex

A. B. Sereno*† & J. H. R. Maunsell*‡

* Division of Neuroscience S-603 and ‡ Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

† Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, New Jersey 07102, USA

The extrastriate visual cortex can be divided into functionally distinct temporal and parietal regions, which have been implicated in feature-related ('what') and spatial ('where') vision, respectively¹. Neuropsychological studies of patients with damage to either the temporal or the parietal regions provide support for this functional distinction^{2–4}. Given the prevailing modular theoretical framework and the fact that prefrontal cortex receives inputs from both temporal and parietal streams^{5,6}, recent

studies have focused on the role of prefrontal cortex in understanding where and how information about object identity is integrated with (or remains segregated from) information about object location^{7–10}. Here we show that many neurons in primate posterior parietal cortex (the 'where' pathway) show sensory shape selectivities to simple, two-dimensional geometric shapes while the animal performs a simple fixation task. In a delayed match-to-sample paradigm, many neuronal units also show significant differences in delay-period activity, and these differences depend on the shape of the sample. These results indicate that units in posterior parietal cortex contribute to attending to and remembering shape features in a way that is independent of eye movements, reaching, or object manipulation. These units show shape selectivity equivalent to any shown in the ventral pathway.

Previous studies of the parietal cortex that demonstrated its sensitivity to object shape have tended to focus on tasks involving hand manipulation of three-dimensional solid objects^{11–14}. Little attention, however, has been paid to the basic question of simple, two-dimensional shape selectivity in the parietal cortex. Here we test directly the extent to which units in the lateral intraparietal area (LIP) in posterior parietal cortex respond to differently shaped, two-dimensional visual stimuli that the animal did not and could not manipulate.

We recorded from 124 isolated neurons in area LIP of two macaque monkeys. The monkeys were trained to perform a simple fixation task. After the animal fixated a central spot, a shape was briefly presented within the receptive field of the unit being recorded. We recorded the activity of 74 of the 124 neurons while the animal performed this fixation task. Surprisingly, many units (42 of 74, 57%) showed a significant difference in activity during the stimulus presentation, which was dependent on which of eight shapes was presented (Fig. 1).

We calculated a shape-selectivity index (SI) for each unit using the average rate of firing for the stimuli that produced the strongest and weakest responses ($SI = (\max - \min)/(\max + \min)$). The histogram in Fig. 2 shows the distribution of indices, with units that show significant differences in responses to different shapes being indicated in black. The median of the indices for these significant units corresponded to a response that was 2.5 times stronger for the most-preferred (best) relative to the least-preferred (worst) shape (median SI = 0.43).

This shape selectivity is unlikely to arise from accidental interactions between shape features and receptive-field profiles, because LIP receptive fields are typically large and homogeneous¹⁵. Nevertheless, to discount such accidental interactions, we tested some units using stimuli in different positions or of different sizes. For 6 of the 42 units with significant differences in response that were dependent on the shape of the stimulus, the same stimuli were presented at the same eccentricity in a second location. The polar angular difference between the first and second location ranged from 25 to 130 degrees depending on the size of the receptive field of the particular unit. There was good agreement between the shape selectivities in the two locations. For three of the units, the preferred shape or preferred two shapes in one location were the same in the second location. For the remaining three units, which were more broadly tuned, at least one or two of the three shapes with the strongest response agreed. In addition, the least-preferred one or two shapes in one location remained the same in the second location. For two units with significant sensitivity to shape, we recorded the activity of each unit when the size of the stimuli was increased by 50%. Each unit maintained its shape preference. These results indicate that shape selectivities in LIP units are not an accidental result of receptive-field profiles.

To test for shape-selective behavioural effects in area LIP, we trained the animals on a delayed match-to-sample task. After the monkey fixated a central spot, a sample shape was briefly presented in one of three eccentric locations. After a short delay (0.5–2.1 s),

three test shapes appeared simultaneously. The animal had to make a saccade to the test shape that matched the sample shape, regardless of sample and test-stimulus locations. Many units (41 of 124 units, 33%) showed a significant difference in activity during the delay period depending on the shape of the sample (Fig. 3). Such selective activation probably contributes to the animal's ability to attend to and remember the sample shape during the period in which no stimulus is on the screen.

For the most responsive location of each unit during the delay

period, we calculated a shape-selectivity index by dividing the average spike-rate difference of the shape with the greatest delay activity (max) and the shape with the weakest delay activity (min) by the sum of their average spike rates. The histogram in Fig. 4 represents the population of units. Units with significant shape-dependent differences in delay activity are indicated in black. These units showed a 73% increase in firing rate for the most-preferred (best) shape compared with the least-preferred (worst) shape (median SI = 0.27).

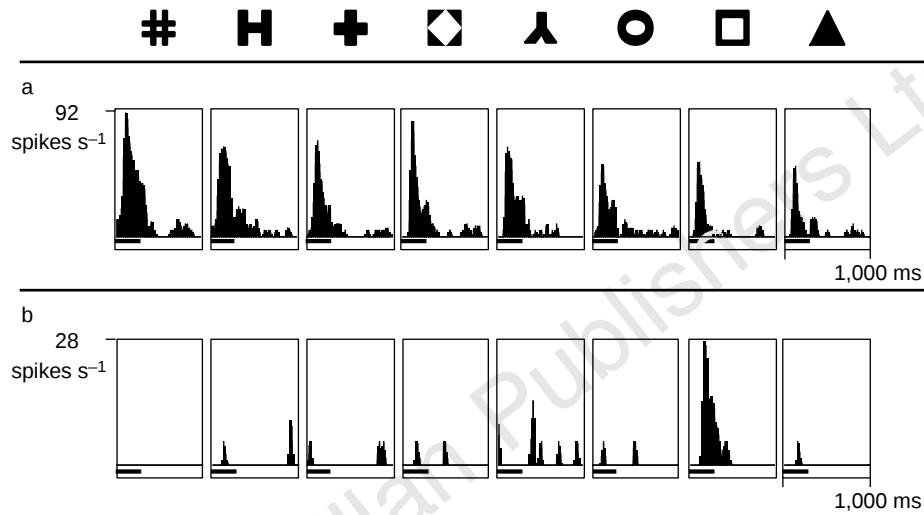


Figure 1 Responses of two neurons during the passive fixation task. **a**, Broadly tuned, shape-selective sensory response (SI = 0.40). **b**, Narrowly tuned, shape-selective sensory response (SI = 0.83). Peristimulus-time histograms from each of two neurons for all eight stimulus conditions are presented with the corresponding shape indicated above its respective histograms. Each histogram

shows the activity of the neuron during the stimulus period (horizontal bar) and the fixation period that followed for each of six trials (**a**) or five trials (**b**). For the unit in **a**, the histograms corresponding to the different shape conditions are ordered by decreasing magnitude of activity during the sample stimulus period.

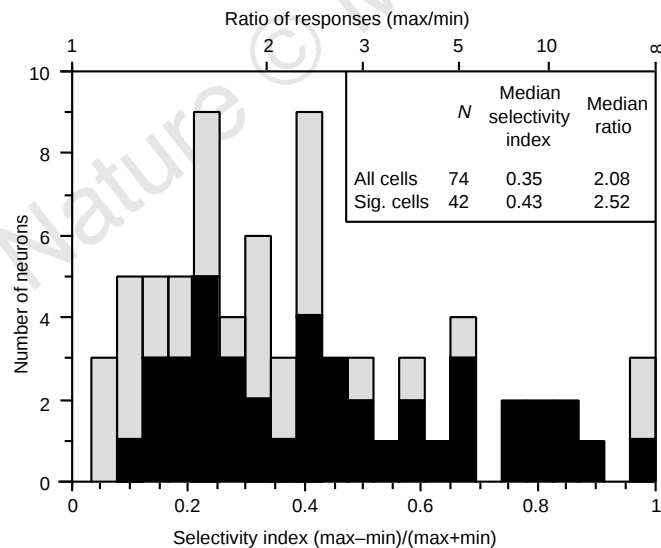


Figure 2 Histogram showing the magnitude of the sensory shape selectivity during the passive fixation task the population of units recorded. Units with significant shape-dependent differences in sample activity are indicated in black. The median selectivity index for the significant units was 0.43.

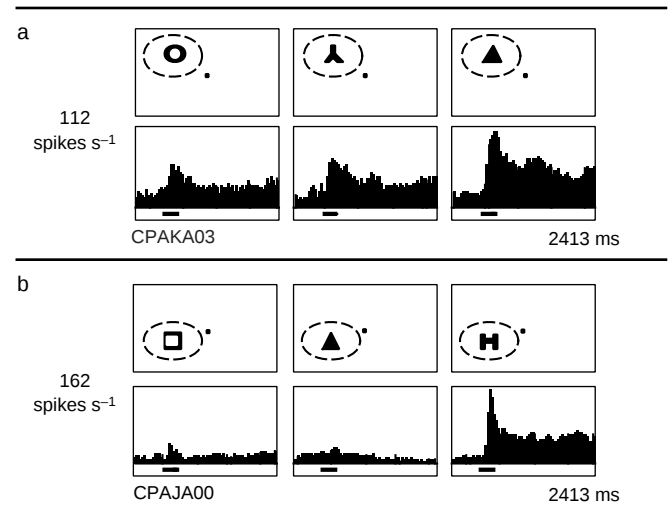


Figure 3 Responses of two neurons during the delayed match-to-sample task. **a**, Broadly tuned, shape-selective delay-period activity. **b**, Narrowly tuned, shape-selective, delay-period activity. Peristimulus-time histograms from a neuron for three particular sample stimulus conditions are presented. The dashed oval represents the neuron's receptive field. Each histogram was compiled from 12 trials and shows the activity of the neuron during the presample fixation period, the sample period (horizontal bar), and the delay period. These neurons show shape selectivity during both the sample and the delay periods. Although only the fixation spot (small black square) was present during the delay period for all three conditions, the neuron in (**b**) was much more active when the sample H-shape had been presented inside its receptive field.

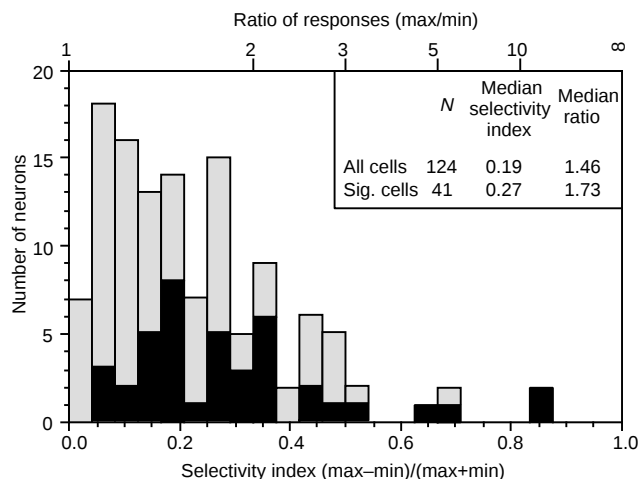


Figure 4 Histogram showing the magnitude of the delay-period shape selectivity during the match-to-sample task for the population of units recorded. Units with significant shape-dependent differences in delay-period activity are indicated in black. The median selectivity index for the significant units was 0.27. Although, for many units, we attempted to include the shapes that seemed to produce the best and worst response among the three shapes used in the task, it is possible that, compared with the sensory shape effects reported above using all eight shapes, we may have underestimated the delay-period shape effects.

Although LIP is important for spatial perception, it represents a late stage in visual processing and has strong associations with structures involved in the planning and execution of eye movements¹⁶. Thus there is much debate over whether units in LIP, and in posterior parietal cortex more generally, represent the spatial whereabouts of objects, or whether they guide parts of the body to the location that these objects occupy^{17–19}. The latter view, which takes into account the output requirements of the ‘where’ pathway, has been put forward to explain shape selectivity in the ‘where’ pathway. For example, neurons in the anterior intraparietal area can be sensitive to those visual qualities of an object that determine the posture of the hand and fingers during a grasping movement^{20–22}.

However, our results present problems for a strictly intentional view of the function of posterior parietal cortex²³. Many LIP units showed changes in response to different shapes presented in their receptive field that could not be accounted for by different eye movements. Most units also showed shape selectivities even in a passive fixation task. The shape selectivity we report in LIP presents an enigma. Because different shapes are presented in the same location, it is difficult to see how they entail different eye-movement programming, whether or not the animal makes an eye movement. Moreover, shape selectivity in LIP is consistent with anatomical studies showing that LIP receives projections from areas in the ‘what’ pathway, including areas V4, TEO and TE (ref. 6). Neurophysiological studies also indicate that the temporal waveforms of neuronal discharge in LIP may contain information that can be used to discriminate stimulus patterns^{24,25}. Shape selectivity in the ‘where’ pathway is a relatively unexplored frontier that promises to change our understanding of cortical visual processing. □

Methods

Tasks and procedures. During the passive fixation task, the stimulus shape was positioned so that it fell within the receptive field. The stimulus for each trial was selected from a set of eight different shapes. Each was a simple two-dimensional black-and-white geometric form (Fig. 1). Each shape fit within the same-sized square region and had an equal number of bright pixels. All eight shapes were centred on the same position within the unit’s receptive field. In each trial, the randomly selected shape was presented four times before a central fixation spot was extinguished. Each repetition of the shape was presented for a

constant duration, ranging from 250 to 300 ms among the units recorded. Likewise, each interstimulus interval was of constant duration, ranging from 500 to 750 ms. The animals were required to maintain fixation within 0.5° of the central 0.1° spot in the centre of the video display throughout the trial. Eye position was monitored using the scleral search-coil method. The animals were rewarded for maintaining fixation on the central spot until it disappeared. For most of the units recorded, stimulus eccentricity ranged between 3.5° and 6.0° and stimulus size ranged between 0.4° and 0.8°.

During the delayed match-to-sample task, after the animal fixated a 0.1° spot in the centre of the video display, a sample shape was presented (ranging from 200 to 400 ms among the units recorded) in one of three locations. The sample was immediately followed by a full-screen pattern mask (duration 15 ms), followed by a variable delay period, ranging from 500 to 2,100 ms among the units recorded. Finally, a test array of three shapes appeared, equidistant from the fixation spot. The animal was required to make an eye movement to the test shape that matched the sample shape for a juice reward. For each unit, samples and test stimuli were a subset of three of eight different shapes (Fig. 1). For most units, the two shapes that elicited the strongest response and the one shape that elicited the weakest response were selected. In catch trials (up to 20% of trials), the test array did not include the sample shape, which was replaced instead by one of the five other shapes. In these trials, the animal was rewarded for maintaining fixation on the central spot until it disappeared (an additional 1,300–1,900 ms). The stimulus arrays were configured so that at least one stimulus position fell within the receptive field. For most of the units recorded, stimulus eccentricity ranged between 3.5° and 6.0° and stimulus size ranged between 0.4° and 0.8°.

Histology. In one animal, histological reconstruction showed, on the basis of myeloarchitectonic criteria^{16,26}, that the 85 units recorded in the lateral bank of the intraparietal sulcus lay within area LIP.

Data analysis. For statistical analysis in the passive fixation task, an *F*-test for shape selectivity was performed on the average rate of firing for the eight different shapes during the sample period, starting 50 ms poststimulus (collapsed across repetitions within as well as across trials). For statistical analysis in the delayed match-to-sample task, an *F*-test was performed on the average rate of firing during the last 300 ms of the delay period. For all statistical tests, a significant criterion level of *P* < 0.05 was used.

Received 1 June; accepted 23 July 1998.

1. Ungerleider, L. G. & Mishkin, M. in *The Analysis of Visual Behavior* (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–586 (MIT Press, Cambridge, MA, 1982).
2. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25 (1992).
3. Farah, M. J. *Visual Agnosia* (MIT Press, Cambridge, MA, 1990).
4. Perenin, M. & Vighetto, A. Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain* **111**, 643–674 (1988).
5. Cavada, C. & Goldman-Rakic, P. S. Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J. Comp. Neurol.* **287**, 422–445 (1989).
6. Webster, M. J., Bachevalier, J. & Ungerleider, L. G. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque. *Cerebral Cortex* **5**, 470–483 (1994).
7. Wilson, F. A. W., O. Scialidhe, S. P. & Goldman-Rakic, P. S. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958 (1993).
8. Rao, C. S., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
9. O. Scialidhe, S. P., Wilson, F. A. W. & Goldman-Rakic, P. S. Areal segregation of face-processing neurons in prefrontal cortex. *Science* **278**, 1135–1138 (1997).
10. Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G. & Haxby, J. V. An area specialized for spatial working memory in human frontal cortex. *Science* **279**, 1347–1351 (1998).
11. Taira, M., Mine, S., Georgopoulos, A. P., Murata, A. & Sakata, H. Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp. Brain Res.* **83**, 29–36 (1990).
12. Murata, A., Gallese, V., Kaseda, M. & Sakata, H. Parietal neurons related to memory-guided hand manipulation. *J. Neurophysiol.* **75**, 2180–2186 (1996).
13. Jeannerod, M. The formation of finger grip during prehension. A cortically mediated visuomotor pattern. *Behav. Brain Res.* **19**, 99–116 (1986).
14. Gallese, V., Murata, A., Kaseda, M., Niki, N. & Sakata, H. Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport* **5**, 1525–1529 (1994).
15. Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area II. Spatial properties. *J. Neurophysiol.* **66**, 1109–1124 (1991).
16. Andersen, R. A., Asanuma, C., Essick, G. & Seigel, R. M. Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *J. Comp. Neurol.* **296**, 65–113 (1990).
17. Shadlen, M. Look but don’t touch, or vice versa. *Nature* **386**, 122–123 (1997).
18. Rizzolatti, G., Riggio, L. & Sheliga, B. M. in *Attention and Performance XV* (eds Umiltà, C. & Moscovitch, M.) 231–265 (MIT Press, Cambridge, MA, 1994).
19. Colby, C. L. Action-oriented spatial reference frames in cortex. *Neuron* **20**, 15–24 (1998).
20. Goodale, M. A. et al. Separate neural pathways for the visual analysis of object shape in perception and prehension. *Curr. Biol.* **4**, 604–610 (1994).
21. Jeannerod, M., Arbib, M. A., Rizzolatti, G. & Sakata, H. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* **18**, 314–320 (1995).

22. Sakata, H., Taira, M., Murata, A. & Mine, S. Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cerebr. Cortex* **5**, 429–438 (1995).
23. Snyder, L. H., Batista, A. P. & Andersen, R. A. Coding of intention in the posterior parietal cortex. *Nature* **386**, 167–170 (1997).
24. Goldberg, M. E. & Gottlieb, J. Neurons in monkey LIP transmit information about stimulus pattern in the temporal waveform of their discharge. *Soc. Neurosci. Abstr.* **23**, 17 (1997).
25. Troschianko, T. et al. Human colour discrimination based on a non-parvocellular pathway. *Curr. Biol.* **6**, 200–210 (1996).
26. Blatt, G. J., Andersen, R. A. & Stoner, G. R. Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J. Comp. Neurol.* **299**, 421–445 (1990).

Acknowledgements. We thank K. Briand, R. Klein, S. Lehy and S. O. Scallidhe for comments on the manuscript. This work was supported by awards from the McDonnell–Pew Foundation, NARSAD, NIMH, and NEI. J.H.R.M. is an Investigator with the Howard Hughes Medical Institute. Animal experiments were conducted in accordance with the Baylor College of Medicine and Rutgers University Animal Care Committees.

Correspondence and requests for materials should be addressed to A.B.S. (e-mail: sereno@cmbn.rutgers.edu).

Mechanism of calcium gating in small-conductance calcium-activated potassium channels

X.-M. Xia*, B. Fakler†, A. Rivard*, G. Wayman*, T. Johnson-Pais*, J. E. Keen*, T. Ishii*, B. Hirschberg*, C. T. Bond*, S. Lutsenko‡, J. Maylie§ & J. P. Adelman*

* Vollum Institute, Departments of † Biochemistry and Molecular Biology, and § Obstetrics and Gynecology, Oregon Health Sciences University, Portland, Oregon 97201, USA

† Department of Physiology, University of Tuebingen, Tuebingen, Germany

The slow afterhyperpolarization that follows an action potential is generated by the activation of small-conductance calcium-activated potassium channels (SK channels). The slow afterhyperpolarization limits the firing frequency of repetitive action potentials (spike-frequency adaption) and is essential for normal neurotransmission^{1–3}. SK channels are voltage-independent and activated by submicromolar concentrations of intracellular calcium¹. They are high-affinity calcium sensors that transduce fluctuations in intracellular calcium concentrations into changes in membrane potential. Here we study the mechanism of calcium gating and find that SK channels are not gated by calcium binding directly to the channel α -subunits. Instead, the functional SK channels are heteromeric complexes with calmodulin, which is constitutively associated with the α -subunits in a calcium-independent manner. Our data support a model in which calcium gating of SK channels is mediated by binding of calcium to calmodulin and subsequent conformational alterations in the channel protein.

Calcium ions are ubiquitous regulators of many cellular processes, and eukaryotic cells have evolved a complex system of transmembrane molecules, channels, pumps and exchangers, which maintain intracellular Ca^{2+} concentrations at very low levels, 10–100 nM. This allows rapid metabolic responses to Ca^{2+} fluxes^{4,5}. Several classes of ion channels, such as Ca^{2+} -activated potassium channels, are gated by intracellular Ca^{2+} ions, thereby coupling intracellular Ca^{2+} levels and membrane potential.

The genes encoding a family of SK channels have been cloned. The amino-acid sequences of the three known members of the SK-channel family, SK1, SK2 and SK3, show that the channels share high overall structural homology, with little similarity to members of other potassium-channel subfamilies⁶. Because of the fundamental role played by SK channels in regulating neuronal excitability, we studied the mechanism underlying Ca^{2+} gating of SK channels.

Current recordings from giant inside-out patches of *Xenopus* oocytes showed that all SK-channel subtypes exhibit similar Ca^{2+}

dose–response relationships, with Ca^{2+} concentrations required for half-maximal activation ($K_{0.5}$) of $\sim 0.3 \mu\text{M}$ and a Hill coefficient of ~ 4 (Fig. 1a). Fast piezo-driven application of Ca^{2+} ($10 \mu\text{M}$) showed that onset of current commences within 1 ms, with time constants of activation of 5–15 ms (Fig. 1b–d). This rate is similar to that for the rapid activation of ligand-gated ion channels^{7,8}. SK currents could be repeatedly activated in continuously perfused patches for as long as the patches remained intact without changes in the activation kinetics, and application of protein kinase inhibitors (W7 or H89) or phosphatase inhibitors (okadaic acid) did not affect Ca^{2+} gating (results not shown). ATP and other nucleotides were not present in the intracellular solution. These results indicate that no diffusible second messengers or protein kinases are necessary for SK-channel gating, and that gating reflects interactions between the channel and Ca^{2+} only.

The similarity in the nature of Ca^{2+} gating of the three SK channels indicates that the structural elements underlying Ca^{2+} gating are conserved. The submicromolar affinity for Ca^{2+} is reminiscent of the affinity for Ca^{2+} of an EF-hand Ca^{2+} -binding protein⁹, yet no such motif is present in the SK channels, nor are C2 domains¹⁰ or motifs similar to the ‘calcium bowl’ present in BK channels¹¹. However, if Ca^{2+} ions are directly chelated by the channel protein then, as for other known Ca^{2+} -binding motifs, negatively charged residues, glutamate and aspartate, are likely to mediate Ca^{2+} binding. Therefore, each of the 21 conserved negatively charged residues within the predicted intracellular domains of the SK-channel α -subunits (Fig. 2a) was mutated individually into a neutral amino acid (E $\rightarrow$ Q, D $\rightarrow$ N) in the SK2 channel. Dose–response experiments showed that in no case was Ca^{2+} gating markedly altered (Fig. 2b). In addition, we made two mutant subunits, one containing neutralizations of the four conserved negatively charged residues in the loop connecting transmembrane domains 2 and 3 (2–3 loop), and the other containing neutralizations of the first nine conserved negatively charged residues in the carboxy terminus. Expression of the 2–3-loop mutant resulted in normal, Ca^{2+} -dependent gating, whereas expression of the C-terminal mutant did not give rise to functional channels (Fig. 2c,d), indicating that the proximal part of the C terminus is

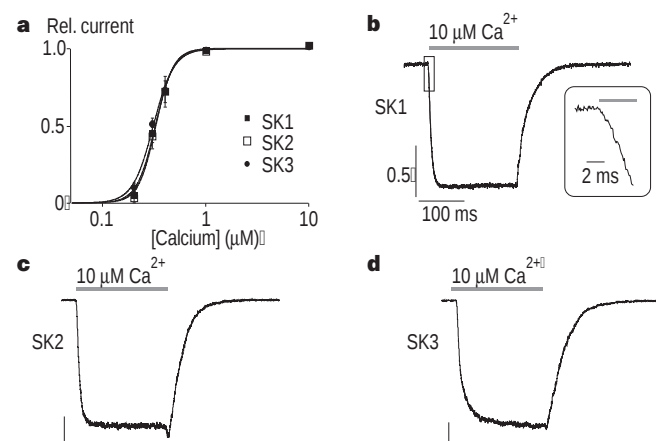


Figure 1 Ca^{2+} gating in SK channels. **a**, Ca^{2+} dose–response relationship for SK1, SK2 and SK3 channels. Relative current amplitude measured at -100 mV is plotted as a function of Ca^{2+} concentration. The data were fitted with the Hill equation, yielding $K_{0.5}$ and Hill coefficient of $0.31 \pm 0.01 \mu\text{M}$ and 4.4 ± 0.2 , $0.33 \mu\text{M} \pm 0.01$ and 5.3 ± 0.9 , and $0.32 \pm 0.03 \mu\text{M}$ and 5.0 ± 0.6 for SK1, 2 and 3, respectively. **b–d**, Fast piezo-driven application of Ca^{2+} ($10 \mu\text{M}$) to inside-out patches expressing SK1 (**b**), SK2 (**c**), or SK3 (**d**) channels. The holding potential was -80 mV ; current and time calibrations are 0.5 nA and 100 ms . Inset for SK1 shows a current increase within 1 ms after Ca^{2+} application. Time constants for activation and deactivation, determined from mono-exponential fits, were 5.8 , 6.3 and 12.9 ms for activation and 21.7 , 29.6 and 38.1 ms for deactivation of SK1, SK2 and SK3, respectively.

important for gating. Introducing a stop codon at position 497 or 477 (Fig. 2a) resulted in the formation of channels that were still gated by Ca^{2+} . However, further truncations at positions 463, 444 or 421 resulted in loss of channel function. Analysis of this region, between the cytoplasmic border of transmembrane domain 6 and position 497, showed that it is the most conserved intracellular domain in the SK channels and may be modelled as a series of four α -helices (A–D boxes, Fig. 3). These results indicate either that Ca^{2+} is bound to this region through a novel Ca^{2+} -binding motif, or that another protein may interact with the C-terminal domain and mediate Ca^{2+} gating.

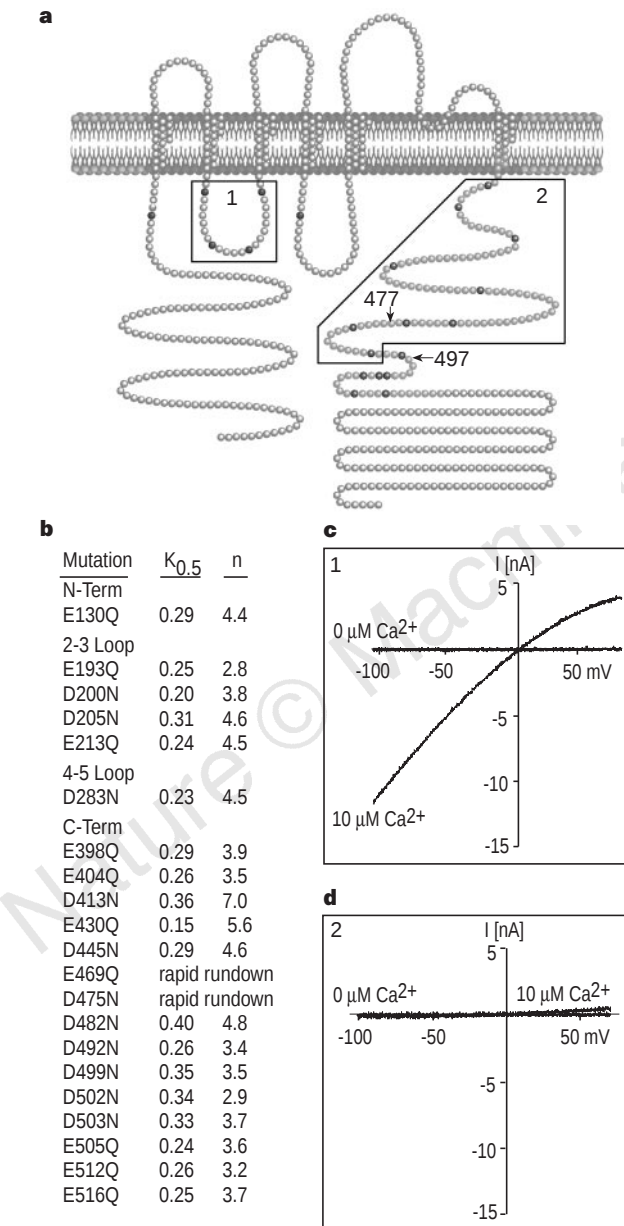


Figure 2 Domains involved in Ca^{2+} gating. **a**, The SK2 subunit. Each circle represents a single residue, and the positions of the conserved intracellular negatively charged amino acids are shaded darkly. The regions in which multiple charge neutralizations were introduced, that is, the 2–3 loop and the proximal domain of the intracellular C terminus, are boxed, and the positions of the truncations are indicated by arrows. **b**, $K_{0.5}$ and Hill coefficient, n , as determined by fitting the Hill equation to the experimentally determined Ca^{2+} dose–response curves for each mutant. **c–d**, Current responses to voltage ramps from -100 mV to 70 mV (2-s duration) in the presence and absence of $10\text{ }\mu\text{M}$ Ca^{2+} from patches expressing the 2–3-loop (**c**) or C-terminus (**d**) mutations.

If Ca^{2+} gating is mediated by association of the channel α -subunits with an endogenous Ca^{2+} -binding β -subunit, then the β -subunit must be expressed at high levels and must bind Ca^{2+} with an affinity similar to that of an EF-hand Ca^{2+} -binding protein. One protein that meets these requirements is calmodulin. Therefore, we used the yeast two-hybrid system to determine whether regions of the C-terminal domain of the SK channels interact with calmodulin. The region consisting of helices A to D from each of the SK channels complemented calmodulin (Fig. 3). Further experiments using subdomains of SK2 showed that helices A–C were sufficient for complementation, whereas regions B–C, B–D, the intracellular amino terminus, the 2–3 loop, the 4–5 loop, and the remainder of the C terminus following the D box (SK2postD) did not interact with calmodulin in this assay (Fig. 3).

We confirmed the interaction between SK2 and calmodulin by glutathione S-transferase (GST)-pulldown and biotinylated-calmodulin-overlay experiments. Fusion proteins containing GST and region A–D, B–D, B–C, SK2post D or the intracellular C terminus of the voltage-gated potassium channel Kv1.1 were bound to glutathione–agarose beads and exposed to purified bovine brain calmodulin, in either the absence or the presence of $10\text{ }\mu\text{M}$ Ca^{2+} . After extensive washes, the retained proteins were eluted with reduced glutathione and prepared as a western blot. The blots were probed with a monoclonal anti-calmodulin antibody, and the results showed that in the absence of Ca^{2+} only regions A–D bound calmodulin, whereas in the presence of Ca^{2+} regions B–C and B–D were also able to bind calmodulin (Fig. 4a, left). Region A–D also retained calmodulin when exposed to protein extracts prepared from *Xenopus* oocytes or rat brain (Fig. 4a, right). Probing western blots of the different subdomain fusion proteins with biotinylated calmodulin also showed that biotinylated calmodulin was bound to region A–D in the absence or presence of Ca^{2+} ,

2-Hybrid analysis

Channel domain		CaM
SK1 (aa 279–425)	A–D	●
SK2 (aa 390–533)	A–D	●
SK3 (aa 539–682)	A–D	●
SK2 (aa 423–487)	B–C	
SK2 (aa 423–533)	B–D	
SK2 (aa 390–487)	A–C	●
SK2 (aa 534–580)	post D	
SK2 (aa 1–80)	N-terminus	
SK2 (aa 81–134)	N-terminus	
SK2 (aa 180–214)	2–3 Loop	
SK2 (aa 279–306)	4–5 Loop	
SK2	A–D –/N	

Figure 3 Calmodulin interacts with SK channels in a yeast two-hybrid assay. Yeast two-hybrid analysis of the interactions between calmodulin and region A–D of SK1, SK2, or SK3, or subdomains of SK2. Viable colonies indicate interaction between calmodulin (CaM) and region A–D from each of the SK channels. aa, amino acids.

whereas binding to region B–C or B–D was Ca^{2+} -dependent (Fig. 4b). The association between SK channels and calmodulin was further demonstrated by immunoprecipitations. We used a rabbit polyclonal antibody to the intracellular domain of SK3 for immunoprecipitations with rat brain extracts. Immunoprecipitated proteins were prepared as a western blot and probed with the anti-calmodulin monoclonal antibody. Calmodulin was detected in the immunoprecipitate produced using the anti-SK3 antibody but not from immunoprecipitates using an anti-Myf5 antibody (Fig. 4c) or several other rabbit polyclonal antibodies (results not shown). The Ca^{2+} independence of calmodulin binding to region A–D, like its binding to phosphorylase kinase¹², indicates that calmodulin binds constitutively to SK channels.

We performed patch-clamp experiments and attempted to disrupt gating by interfering with the association between the SK α -subunits and calmodulin. Application of 10 μM calmidazolium¹³ or

10 μM of a peptide inhibitor of calmodulin¹⁴ to inside-out patches containing SK2 channels had no effect on current amplitudes evoked with 1 μM Ca^{2+} . Similarly, efforts to strip calmodulin away from the channels by altering ionic strength (with 5–500 mM potassium gluconate) or pH (pH 5.0–9.5) were ineffective (results not shown). These results indicate that the binding between calmodulin and SK channels is similar to the binding of calmodulin to phosphorylase kinase, for which denaturing conditions and EGTA are necessary to dissociate calmodulin from the enzyme complex¹⁵. The results also support the conclusion that other calmodulin-binding proteins, such as calmodulin kinases or calcineurin, are not involved directly in SK-channel gating, as their activities would have been affected by calmidazolium or the calmodulin-inhibitory peptide.

Mutagenesis of aspartate to alanine in the first position of the EF-hand motif alters the affinity of calmodulin for Ca^{2+} (refs 16, 17). Therefore, we made three mutant rat calmodulins, one with the D $\rightarrow$ A mutation in the third and fourth EF hands, calmodulin($\text{D}_{\text{EF}3,4}\text{A}$), one with this mutation in the second, third and fourth EF hands, calmodulin($\text{D}_{\text{EF}2,3,4}\text{A}$), and one with the mutation in all four EF hands, calmodulin($\text{D}_{\text{EF}1,2,3,4}\text{A}$). We expressed the wild-type rat calmodulin, which is identical to *Xenopus* calmodulin¹⁸, calmodulin($\text{D}_{\text{EF}3,4}\text{A}$), calmodulin($\text{D}_{\text{EF}2,3,4}\text{A}$) and calmodulin($\text{D}_{\text{EF}1,2,3,4}\text{A}$) in *Xenopus* oocytes and loaded the extracted proteins on a gel for western blot analysis in the presence or absence of Ca^{2+} ; we probed the blot with the anti-calmodulin antibody. Under each condition, we detected a single band for wild-type calmodulins, which showed a Ca^{2+} -dependent mobility shift (Fig. 5a). In contrast, we detected

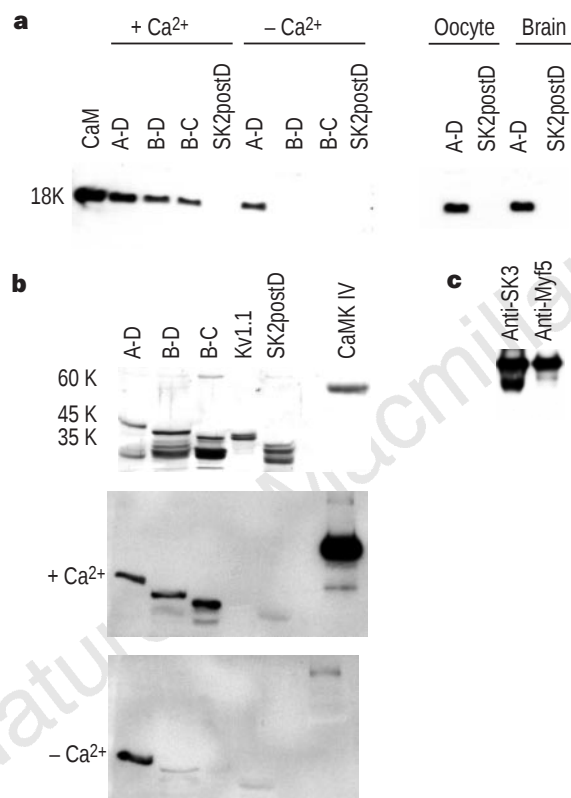


Figure 4 Calmodulin binds to SK channels. **a**, Left, the indicated SK2 fusion proteins were bound to glutathione-agarose and incubated with purified bovine brain calmodulin (CaM). Bound proteins were eluted and prepared as a western blot and were probed with a calmodulin-specific monoclonal antibody. Experiments in lanes 2–5 were done in the presence of 10 μM Ca^{2+} ; experiments in lanes 6–9 were done in the absence of Ca^{2+} (presence of 5 mM EGTA). Right, a fusion protein containing GST and region A–D of SK2 specifically retains calmodulin from oocyte extracts or rat brain. An excess of protein from oocytes or rat brain was incubated with glutathione-agarose-bound GST/A–D fusion protein. **b**, Top, Coomassie-stained gel showing the indicated GST-fusion proteins and calmodulin kinase IV (CaMK IV). Middle, biontynylated calmodulin overlay performed in the presence of 10 μM Ca^{2+} . Bottom, biontynylated calmodulin overlay performed in the absence of Ca^{2+} (presence of 5 mM EGTA). **c**, Co-immunoprecipitation of calmodulin with anti-SK3 antibody. Protein extract (100 μg) from rat brain was immunoprecipitated with a rabbit polyclonal antibody against the intracellular C terminus of SK3, including the conserved A–D region, or with a rabbit polyclonal antibody against Myf5. Immunoprecipitates were prepared as a western blot and probed with the anti-calmodulin monoclonal antibody. The cross-reacting band seen in both lanes at apparent relative molecular mass $\sim 25\text{K}$ is antibody light chain.

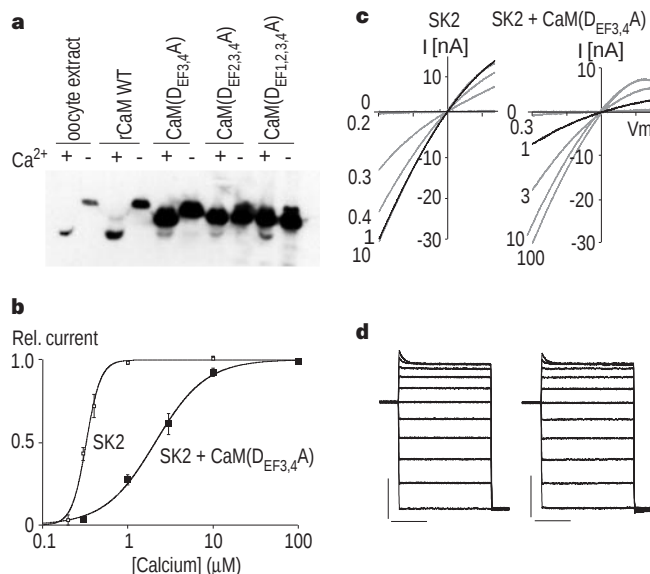


Figure 5 Calmodulin mediates Ca^{2+} gating of SK channels. **a**, Western blot prepared from oocytes expressing the indicated calmodulin (CaM) mutants. Samples were electrophoresed in loading buffer containing either 5 mM Ca^{2+} (+) or 2 mM EGTA (–), and the blot was probed with the anti-calmodulin antibody. The same amount of protein was loaded in each lane. r, rat; WT, wild-type. **b**, Ca^{2+} dose-response relation of SK2 channels expressed alone or with $\text{CaM}(\text{D}_{\text{EF}3,4}\text{A})$. Relative current amplitudes measured at -100 mV are plotted as a function of Ca^{2+} concentration (mean $\pm$ s.d. of $n = 6$). Lines represent a fit of the Hill equation to the data, yielding $K_{0.5}$ and a Hill coefficient of 0.32 μM and 5.1 for SK2 and 2.10 μM and 1.5 for SK2 co-expressed with $\text{CaM}(\text{D}_{\text{EF}3,4}\text{A})$. **c**, Current responses to voltage ramps from -100 mV to 70 mV (2-s duration) in the presence of the indicated intracellular Ca^{2+} concentration for SK2 or SK2 coexpressed with $\text{CaM}(\text{D}_{\text{EF}3,4}\text{A})$. Traces recorded in the presence of 1 μM Ca^{2+} are shown by dark lines. **d**, Currents evoked by voltage steps in an inside-out patch expressing SK2 (left) or SK2 with $\text{CaM}(\text{D}_{\text{EF}3,4}\text{A})$ (right) in the presence of 10 μM Ca^{2+} . The voltage steps (50 ms) were from -100 mV to 100 mV in 20-mV increments from a holding potential of 0 mV. The current and time calibrations are 10 nA and 20 ms, respectively.

two bands for calmodulin(D_{EF2,3,4A}) or calmodulin(D_{EF1,2,3,4A}), namely a minor band that probably corresponds to endogenous *Xenopus* calmodulin, which showed the Ca²⁺-dependent migration, and a second band which did not show a Ca²⁺-dependent mobility shift and migrated differently from either of the wild-type calmodulin bands (Fig. 5a). This band probably represents calmodulin(D_{EF2,3,4A}) or calmodulin(D_{EF1,2,3,4A}). Calmodulin(D_{EF3,4A}) migrated between the two positions seen for wild-type calmodulin and showed a partial mobility shift. These results are consistent with previous results¹⁶ and indicate that the mutations affect the Ca²⁺-binding affinity of calmodulin. The results shown in Fig. 5a reflect equal amounts of oocyte extract loaded in each lane, showing that the mutant calmodulins are highly expressed, even compared with expression of wild-type rat calmodulin and native oocyte calmodulin.

We expressed SK2 and calmodulin(D_{EF3,4A}) or calmodulin(D_{EF2,3,4A}) in *Xenopus* oocytes, and evaluated Ca²⁺ gating electrophysiologically. In both cases, the Ca²⁺ dose-response curves showed that the K_{0.5} for Ca²⁺ gating was shifted by more than sixfold and the Hill coefficient was reduced from ~4 to ~2 (Fig. 5b, c and Table 1). In contrast, dose-response curves resulting after expression of SK2 and wild-type rat calmodulin were not different from those resulting after expression of SK2 alone (Table 1). Currents evoked by voltage steps when SK2 was expressed with calmodulin(D_{EF3,4A}) or calmodulin(D_{EF2,3,4A}) were not different from those produced after expression of SK2 alone (Fig. 5d). Expression of SK2 and calmodulin(D_{EF1,2,3,4A}) resulted in a markedly reduced current amplitude (reduced by >50-fold compared with expression of SK2 with wild-type calmodulin) but showed no shift in Ca²⁺ sensitivity compared with expression of SK2 alone (not shown).

These results suggest a model in which the sensitivity of SK-channel gating to Ca²⁺ reflects binding of Ca²⁺ to constitutively SK-associated calmodulin. The biochemical results and the lack of wash-out in inside-out patches indicate that calmodulin is tightly bound within a 'shell' provided by the A-D-box region of the SK α -subunits. However, within this framework, Ca²⁺-dependent conformational shifts take place, as reflected by the Ca²⁺-dependence of calmodulin binding to the B-C or B-D regions (Fig. 4). Conformational changes in calmodulin initiated by Ca²⁺-binding are translated to the channel α -subunits, resulting in gating. This type of intimate association between calmodulin and the α -subunits is compatible with the rapid activation of the channels (Fig. 1b). Thus, calmodulin is an integral component of SK channels that probably does not contribute to the ion-conducting pore. SK channels are, therefore, different from other heteromeric ion channels with several distinct α -subunits^{19–23}, and from those with associated β -subunits that alter gating parameters but are not fundamentally necessary for gating itself (for review, see ref. 24).

Intracellular Ca²⁺ concentrations and membrane potential are important metabolic parameters in all organisms. SK channels have undergone remarkably few structural changes over long evolutionary periods, as shown by the homology between the mammalian SK channels and their counterparts in *Drosophila* and *Caenorhabditis elegans* (our unpublished results). Calmodulin is present in all eukaryotic cells, having evolved very early on as a ubiquitous transducer of intracellular Ca²⁺ signals, and has also undergone minor structural changes only^{18,25}. Indeed, on the basis of analyses of

Paramecium mutants, it was suggested that calmodulin regulates the activity of a Ca²⁺-activated potassium channel²⁶. It is likely that these two components were structurally and functionally fused at early stages of eukaryotic evolution. □

Methods

Molecular biology and electrophysiology. Site-directed mutagenesis was performed as described²⁷. *In vitro* messenger RNA synthesis and oocyte injections were performed as described⁶. Giant-patch recordings were made 3–7 days after injection of proteins. Pipettes made from thick-walled borosilicate glass (A-M Systems), had resistances of 0.3–0.6 M Ω when filled with (in mM) KOH, 116; KCl, 4; HEPES, 10; and CaCl₂, 1.8; pH adjusted to 7.2 with MES. Inside-out patches were superfused with an intracellular solution containing (in mM): KOH, 116; KCl, 4; HEPES, 10; EGTA, 1; this solution was supplemented with CaCl₂, and the pH was adjusted to 7.2 with MES; the amount of CaCl₂ required to yield the concentrations indicated was calculated according to ref. 28. Rapid exchange of Ca²⁺ was achieved using a piezo-driven application system; the time constant of solution exchange was 0.5 ms (ref. 29). Ca²⁺ dose-response curves were obtained as described⁶. All data points are presented as mean $\pm$ s.d. of *n* experiments.

Yeast two-hybrid analysis. The indicated SK-channel coding sequences were subcloned into vector pPC97, as fusions with the GAL4 DNA-binding domain. Rat calmodulin was fused to the transcriptional-activator domain in pPC86 (ref. 30). HF7c yeast were co-transformed with the indicated plasmids; growth was monitored after 2 days of incubation at 30 °C on medium lacking leucine and tryptophan.

GST-fusion proteins and western blots. Channel sequences (A–D, residues 390–534; B–D, residues 423–534; BC, residues 423–488; C terminus, residues 528–580) were subcloned into pGEX-KG (Pharmacia). Bacterial lysates or extracts from *Xenopus* oocytes or rat brain were incubated with glutathione-agarose (Sigma) and subsequently washed in the presence (10 μ M Ca²⁺) or absence (5 mM EGTA) of Ca²⁺. Resin-bound proteins were incubated with purified bovine brain calmodulin (a gift from D. Brickey) in the presence or absence of Ca²⁺, and bound proteins were eluted with reduced glutathione (Sigma). SDS-PAGE was performed with 0.5 mm EGTA in the gel and running buffer and proteins were electroblotted to Hybond membrane (Amersham). Calmodulin was detected using a monoclonal antibody (Upstate Biotechnology) and horseradish-peroxidase-linked secondary antibody (Bio-Rad), visualized with chemiluminescence (NEN).

Biotinylated calmodulin overlays. GST-fusion proteins were prepared as a western blot on Immobilon-P membranes (Millipore) and incubated at room temperature in blocking buffer (1% BSA in 50 mM Tris, pH 7.5, 0.2 M NaCl, 0.5 mM CaCl₂ or 1 mM EGTA, and 50 mM MgCl₂). Biotinylated calmodulin (Calbiochem) was added to 100 ng ml⁻¹, and incubated for 1 h. The membrane was subsequently washed in blocking buffer containing 0.05% TWEEN 20, and incubated with streptavidin-alkaline phosphatase (Calbiochem) in the same buffer, followed by washes with blocking buffer. Bound biotinylated calmodulin was visualized with electrochemiluminescence (NEN).

Co-immunoprecipitation. Protein extracts were incubated with either anti-SK3 polyclonal antibody or anti-Myf-5 polyclonal antibody (C-20) (Santa Cruz Biotechnology). Antigen-antibody complexes were collected with A/G Plus agarose (Santa Cruz Biotechnology). Immunoprecipitated protein was boiled in SDS sample buffer containing 0.5 mM EGTA, and prepared as a western blot. Calmodulin was detected with the anti-calmodulin antibody (Upstate Biotechnology).

Received 12 June; accepted 2 September 1998.

- Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, 1992).
- Lancaster, B., Nicoll, R. A. & Perkel, D. J. Calcium activates two types of potassium channels in rat hippocampal neurons in culture. *J. Neurosci.* **11**, 23–30 (1991).
- Sah, P. Properties of channels mediating the apamin-insensitive afterhyperpolarization in vagal motoneurons. *J. Neurophysiol.* **74**, 1772–1776 (1995).
- Schatzmann, H. J. in *Membrane Transport of Calcium* (ed. Carafoli, E.) 41–108 (Academic, London, 1982).
- Clapham, D. E. Calcium signaling. *Cell* **80**, 259–268 (1995).
- Köhler, M. *et al.* Small-conductance, calcium-activated potassium channels from mammalian brain. *Science* **273**, 1709–1714 (1996).
- Lester, R. A., Clements, J. D., Westbrook, G. L. & Jahr, C. E. Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature* **346**, 565–567 (1990).
- Maconochie, D. J., Zempel, J. M. & Steinbach, J. H. How quickly can GABA_A receptors open? *Neuron* **12**, 61–71 (1994).

Table 1 Parameters characterizing Ca²⁺ gating of SK2 channels

	K _{0.5} (μ M)	Hill	<i>n</i>
SK2	0.33 $\pm$ 0.01	5.3 $\pm$ 0.9	6
SK2 + calmodulin	0.35 $\pm$ 0.04	4.2 $\pm$ 0.4	5
SK2 + calmodulin(D _{EF3,4A})	2.13 $\pm$ 0.33	1.5 $\pm$ 0.1	6
SK2 + calmodulin(D _{EF2,3,4A})	1.97 $\pm$ 0.14	2.2 $\pm$ 0.1	5

The parameters characterizing Ca²⁺ gating of SK2 channels expressed alone or co-expressed with the indicated calmodulins are shown. Values are presented as mean $\pm$ s.d. of *n* experiments.

9. Persechini, A., Moncrief, N. D. & Kretsinger, R. H. The EF-hand family of calcium-modulated proteins. *Trends Neurosci.* **12**, 462–467 (1989).
10. Shao, X. *et al.* Bipartite Ca^{2+} -binding motif in C_2 domains of synaptotagmin and protein kinase C. *Science* **273**, 248–251 (1996).
11. Schreiber, M. & Salkoff, L. A novel calcium-sensing domain in the BK channel. *Biophys. J.* **73**, 1355–1363 (1997).
12. Miyano, O., Kameshita, I. & Fugisawa, H. Purification and characterization of a brain-specific multifunctional calmodulin-dependent protein kinase from rat cerebellum. *J. Biol. Chem.* **267**, 1198–1203 (1992).
13. Fischer, T. H., Campbell, K. P. & White, G. C. An investigation of functional similarities between the sarcoplasmic reticulum and platelet calcium-dependent adenosinetriphosphatases with the inhibitors quercetin and calmidazolium. *Biochemistry* **26**, 8024–8030 (1987).
14. Colbran, R. J., Fong, Y. L., Schworer, C. M. & Soderling, T. R. Regulatory interactions of the calmodulin-binding, inhibitory, and autophosphorylation domains of Ca^{2+} calmodulin-dependent protein kinase II. *J. Biol. Chem.* **263**, 18145–18151 (1988).
15. Pictou, C., Klee, C. B. & Cohen, P. Phosphorylase kinase from rabbit skeletal muscle: identification of the calmodulin-binding subunits. *J. Biol. Chem.* **255**, 553–561 (1980).
16. Geiser, J. R., Tuinen, D. V., Brockerhoff, S. E., Neff, M. M. & Davis, T. N. Can calmodulin function without binding calcium? *Cell* **65**, 949–959 (1991).
17. Putkey, J. A., Sweeney, H. L. & Campbell, S. T. Site-directed mutation of the trigger calcium-binding site in cardiac troponin C. *J. Biol. Chem.* **264**, 12370–12378 (1989).
18. Chien, Y. & Dawid, I. B. Isolation and characterization of calmodulin genes from *Xenopus laevis*. *Mol. Cell. Biol.* **4**, 507–513 (1984).
19. Langosch, D., Thomas, L. & Betz, H. Conserved quaternary structure of ligand-gated ion channels: the postsynaptic glycine receptor is a pentamer. *Proc. Natl Acad. Sci. USA* **85**, 7394–7398 (1988).
20. Verdoorn, T. A., Draguhn, A., Ymer, S., Seeburg, P. H. & Sakmann, B. Functional properties of recombinant rat GABA_A receptors depend upon subunit composition. *Neuron* **4**, 919–928 (1990).
21. Monyer, H. *et al.* Heteromeric NMDA receptors: molecular and functional distinction of subtypes. *Science* **256**, 1217–1221 (1992).
22. Herb, A. *et al.* The KA-2 subunit of excitatory amino acid receptors shows widespread expression in brain and forms ion channels with distantly related subunits. *Neuron* **8**, 775–785 (1992).
23. Galzi, J.-L., Revah, F., Bessis, A. & Changeux, J.-P. Functional architecture of the nicotinic acetylcholine receptor: from electric organ to brain. *Annu. Rev. Pharmacol.* **31**, 37–72 (1991).
24. Adelman, J. P. Proteins that interact with the pore-forming subunits of voltage-gated ion channels. *Curr. Opin. Neurobiol.* **5**, 286–295 (1995).
25. Wylie, D. C. & Vanaman, T. C. In *Calmodulin* (eds Cohen, P. & Klee, C. B.) 1–15 (Elsevier, Amsterdam, 1988).
26. Hinrichsen, R. D., Burgess-Cassler, A., Solvsted, B. C., Hennessey, T. & Kung, C. Restoration by calmodulin of a Ca^{2+} -dependent K^+ current missing in a mutant of *Paramecium*. *Science* **232**, 503–506 (1986).
27. Weiner, M. P. *et al.* Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* **151**, 119–123 (1994).
28. Fabiato, A. & Fabiato, F. Calculator programs for computing the composition of the solutions containing multiple metals and ligands for experiments in skinned muscle cells. *J. Physiol.* **75**, 463–505 (1979).
29. Oliver, D., Hahn, H., Antz, C., Ruppersberg, J. P. & Fakler, B. Interactions of permeant and blocking ions in cloned inward-rectifier K^+ channels. *Biophys. J.* **74**, 2318–2326 (1998).
30. Chevray, P. M. & Nathans, D. Protein interaction cloning in yeast: identification of mammalian proteins that reacted with the leucine zipper of Jun. *Proc. Natl Acad. Sci. USA* **89**, 5789–5793 (1992).

Acknowledgements. We thank T. Soderling and P. Ruppersberg for encouragement and fruitful conversations; K. Gibson for site-directed mutagenesis; D. Oliver for technical assistance; and L. Cordelia and E. Wiltshire for artwork and patience. This work was funded by NIH grants and a grant from ICAgent.

Correspondence and requests for materials should be addressed to J.P.A. (e-mail: adelman@ohsu.edu).

Replication checkpoint requires phosphorylation of the phosphatase Cdc25 by Cds1 or Chk1

Yan Zeng^{††}, Kristi Chrispell Forbes^{††}, Zhiqi Wu[§], Sergio Moreno^{||}, Helen Piwnicka-Worms[§] & Tamar Enoch[‡]

^{*} Department of Cell Biology and Physiology, and [§] Howard Hughes Medical Institute, Washington University School of Medicine, Box 8228, 660 South Euclid Avenue, St Louis, Missouri 63110, USA

[‡] Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

^{||} Instituto de Microbiología Bioquímica, Departamento de Microbiología y Genética, CSIC/Universidad de Salamanca, 37007 Salamanca, Spain

[†] These authors contributed equally to this work

Checkpoints maintain the order and fidelity of events of the cell cycle by blocking mitosis in response to unreplacated or damaged DNA¹. In most species this is accomplished by preventing activation of the cell-division kinase Cdc2, which regulates entry into mitosis^{2–5}. The Chk1 kinase, an effector of the DNA-damage checkpoint, phosphorylates Cdc25, an activator of Cdc2 (refs 6–

11). Phosphorylation of Cdc25 promotes its binding to 14-3-3 proteins, preventing it from activating Cdc2 (ref. 8). Here we propose that a similar pathway is required for mitotic arrest in the presence of unreplacated DNA (that is, in the replication checkpoint) in fission yeast. We show by mutagenesis that Chk1 functions redundantly with the kinase Cds1 at the replication checkpoint and that both kinases phosphorylate Cdc25 on the same sites, which include serine residues at positions 99, 192 and 359. Mutation of these residues reduces binding of 14-3-3 proteins to Cdc25 *in vitro* and disrupts the replication checkpoint *in vivo*. We conclude that both Cds1 and Chk1 regulate the binding of Cdc25 to 14-3-3 proteins as part of the checkpoint response to unreplacated DNA.

cds1[−] (ref. 12) and *chk1[−]* (refs 6, 7, 11) cells arrest normally in response to the DNA-replication inhibitor hydroxyurea, indicating that these genes are not essential for the checkpoint response to unreplacated DNA. However, a possible function of Chk1 in this checkpoint has been reported in both fission yeast and *Drosophila melanogaster*^{13–17}. To test whether the replication checkpoint could be mediated by both *chk1⁺* and *cds1⁺*, we constructed *cds1[−]chk1[−]* double mutants and examined their response to hydroxyurea, which causes cells to arrest in the S phase of the cell cycle. *chk1[−]cds1[−]* cells completely lacked a replication checkpoint (Fig. 1). More than 80% of the cells in hydroxyurea-treated *chk1[−]cds1[−]* cultures underwent an aberrant mitosis ('cut') instead of S-phase arrest (Fig. 1a, b). The extent and kinetics of 'cut' formation in these cultures is comparable to that observed for the replication-checkpoint-deficient mutant, *hus1[−]* (Fig. 1a)¹⁸. The double-mutant phenotype has been reported and may indicate that the functions of Chk1 and Cds1 overlap^{16,17}. Alternatively, abnormal DNA structures could arise that activate Chk1 in the absence of Cds1 (ref. 16). Modest overexpression of *chk1⁺* can suppress the hydroxyurea sensitivity of *cds1[−]* cells (Fig. 1c), suggesting that *chk1⁺* can perform some *cds1⁺* functions in response to unreplacated DNA.

Chk1 phosphorylates Cdc25 as part of the checkpoint response to DNA damage^{8–10}. To determine whether Cds1 could work by a similar mechanism, we studied phosphorylation of Cdc25 by Cds1

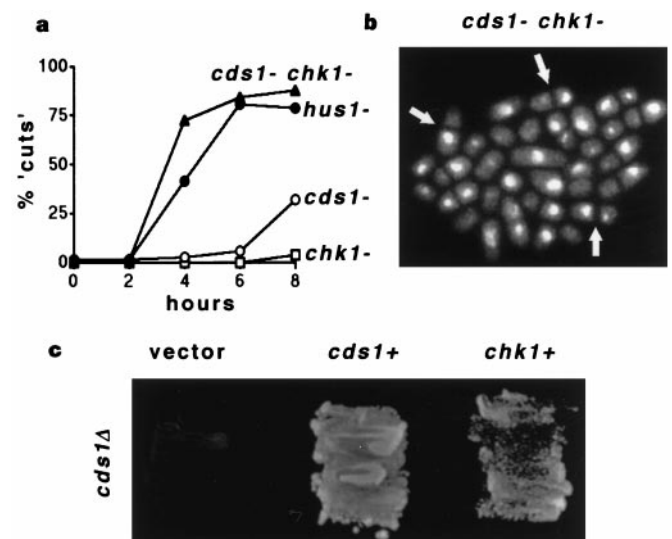


Figure 1 *chk1⁺* and *cds1⁺* function together in the checkpoint response to unreplacated DNA. **a**, Percentage of 'cut' cells in *chk1[−]cds1[−]* (strain TE856), *chk1[−]* (TE790), *cds1[−]* (TE700) and *hus1[−]* (TE481) cultures, incubated for the indicated times in 10 mM hydroxyurea. **b**, Fixed *cds1[−]chk1[−]* cells after 6 h in hydroxyurea. Arrows indicate examples of 'cuts'. **c**, *cds1Δ* (TE700) cells were transformed with vector pREP1 (pTE102), pREP1*chk1⁺* (pTE533), or pREP1*cds1⁺* (pTE535) and grown on minimal plates containing 5 mM hydroxyurea. *chk1⁺* and *cds1⁺* are only moderately overexpressed (see Methods).

and Chk1 *in vitro*. Kinase-active but not kinase-dead forms of Chk1 and Cds1 phosphorylated Cdc25 *in vitro* (Fig. 2a, top). Two-dimensional tryptic phosphopeptide mapping studies were performed to compare the pattern of Cdc25 phosphorylation by Cds1 and by Chk1 (Fig. 2b, c). In each case, at least seven phosphopeptides were resolved. Mixing experiments showed that the phosphopeptides derived from Chk1- and Cds1-phosphorylated Cdc25 co-migrated (Fig. 2d), indicating that Cdc25 is phosphorylated on equivalent sites by Chk1 and Cds1 *in vitro*. Phosphoamino-acid analysis showed that Cdc25 is phosphorylated on serine residues (data not shown).

We identified serines 99 and 359 as sites of phosphorylation by subjecting radiolabelled Cdc25 to proteolysis and analysing the resulting phosphopeptides by reverse-phase high-pressure liquid chromatography (HPLC) followed by Edman degradation (see Methods). We constructed mutations that substituted alanine for serine at these sites, and studied the phosphorylation of mutant proteins by Chk1 and Cds1. Results obtained with Chk1 and Cds1 were identical, so we present only data obtained with Cds1 here. Substitution of alanine for serine at position 99 resulted in the disappearance of phosphopeptide 1 (Fig. 2e), whereas substitution

of alanine for serine at positions 99 and 359 resulted in the loss of phosphopeptides 1, 3 and 5 (Fig. 2f; phosphopeptides 3 and 5 arise from partial proteolysis between R355 and R356). These results establish that serines 99 and 359 are sites of phosphorylation by Chk1 and Cds1 *in vitro*. We identified serine 192 as a minor site of phosphorylation both *in vitro* and *in vivo* (data not shown).

We performed labelling experiments to determine the phosphorylation status of endogenous Cdc25 in the presence and absence of hydroxyurea. A phosphoprotein of relative molecular mass 80,000 (M_r 80K) was detected in anti-Cdc25 immunoprecipitates prepared from the wild-type fission yeast strain but not from a *cdc25* deletion strain (data not shown). The two-dimensional tryptic phosphopeptide maps of Cdc25 radiolabelled in the absence (data not shown) and presence (Fig. 2g) of hydroxyurea were identical. Several phosphopeptides detected *in vitro* were also observed *in vivo* (Fig. 2g, spots 1, 2, 3, 5, 7). Other phosphopeptides (Fig. 2g, spots a–d) were present *in vivo* but were either absent or poorly phosphorylated *in vitro*. To confirm that serines 99, 192 and 359 are phosphorylated *in vivo*, we introduced *cdc25*⁺ and *cdc25*-S3, a mutant containing serine-to-alanine substitutions at positions 99, 192 and 359, into *cdc25* Δ cells. Transformants were labelled and wild-type and mutant Cdc25 phosphoproteins were purified and digested with trypsin. The two-dimensional maps of ectopically expressed Cdc25 (Fig. 2h) were identical to those of endogenous Cdc25 (Fig. 2g) except for the presence of phosphopeptide e. Phosphopeptides 1, 3 and 5 were absent in maps obtained with Cdc25-S3 (Fig. 2i), indicating that these phosphopeptides are due to S99 and S359 phosphorylation. These results establish that Cds1 and Chk1 phosphorylate Cdc25 *in vitro* on a subset of sites that are phosphorylated *in vivo*, and indicate that phosphorylation of these sites should be reduced in a *cds1*[−]*chk1*[−] mutant. The inability to detect differences in Cdc25 phosphorylation in the presence of hydroxyurea may indicate that the replication checkpoint functions to maintain phosphorylation of Cdc25 as it is being stockpiled during the checkpoint response, as is the case for human Cdc25C (ref. 8). Alternatively, activation of the DNA-damage checkpoint by the labelling process may affect our ability to detect phosphorylation differences. Further studies are required to determine whether

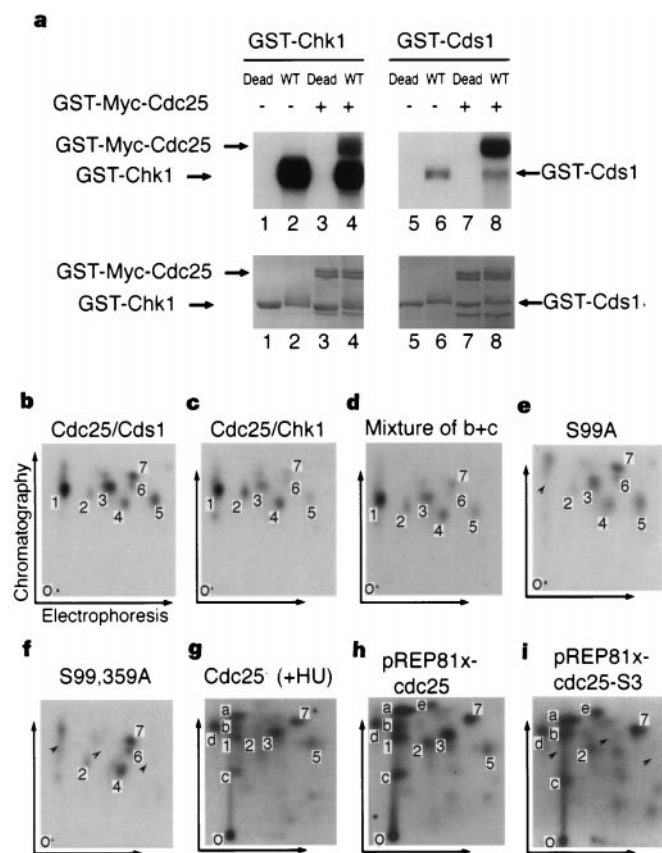


Figure 2 Phosphorylation of Cdc25. **a**, GST-fusions of wild-type (WT) and kinase-inactive (Dead) forms of Chk1 and Cds1 were incubated in the absence or presence of GST-Myc-Cdc25. Kinase assays were performed *in vitro*, resolved by SDS-PAGE visualized by autoradiography (top panel) or Ponceau S staining (bottom panel). **b–i**, Two-dimensional tryptic phosphopeptide mapping of wild-type and mutant Cdc25 phosphorylated *in vitro* (**b–f**) and *in vivo* (**g–i**). **b**, Cdc25 phosphorylated by Cds1. **c**, Cdc25 phosphorylated by Chk1. **d**, A mixture of phosphopeptides from **b** and **c**. **e**, Cdc25(S99A) phosphorylated by Cds1. **f**, Cdc25(S99A,S359A) phosphorylated by Cds1. **g**, Cdc25 from hydroxyurea (HU)-treated wild-type (TE271) cells. **h**, **i**, Cdc25 isolated from *cdc25* Δ cells (TE80) transformed with pREP81 $\times$ *cdc25*⁺ (pTE525) (**h**) or pREP81 $\times$ *cdc25* - S3 (pTE547) (**i**). The origin is depicted by O and arrows depict the location to which missing phosphopeptides would be expected to migrate.

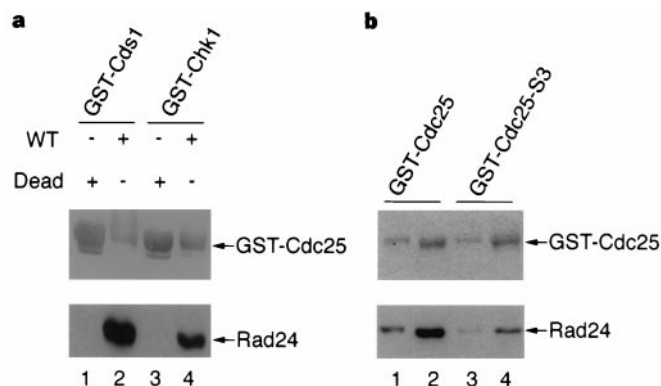


Figure 3 Association between Cdc25 and Rad24 *in vitro*. **a**, Cdc25 bound to GSH beads was incubated with inactive forms of Cds1 (lane 1) or Chk1 (lane 3) or with active forms of Cds1 (lane 2) or Chk1 (lane 4). Washed reactions were incubated with soluble Rad24 and association of Rad24 with Cdc25 was monitored by immunoblotting for Cdc25 (top panel) and Rad24 (bottom panel). **b**, Different concentrations of Cdc25 (lanes 1, 2) and Cdc25-S3 (lanes 3, 4) phosphorylated by Cds1 were incubated with Rad24. Binding was assessed as in **a**.

phosphorylation of serines 99, 192 and 359 is induced or maintained by the replication checkpoint.

The sequences bordering and including serines 99 (RSL_{S99}CT), 359 (RR_{TQS359}MF) and 192 (RS_{RS192}SG) in Cdc25 contain potential recognition motifs for the binding of 14-3-3 proteins¹⁹, although they differ from the consensus motifs in that they contain either threonine, phenylalanine or glycine instead of proline in the +2 position relative to the phosphorylated serine. However, variability in the +2 position is tolerated²⁰. Like human Cdc25C (ref. 8), fission yeast Cdc25 interacted with 14-3-3 proteins when Cdc25 was expressed from viral vectors in insect cells (data not shown). Chk1 phosphorylates fission yeast and human Cdc25 proteins^{8–10} and phosphorylation promotes binding of 14-3-3 protein to human Cdc25C (ref. 8). We tested whether phosphorylation of fission yeast Cdc25 by Cds1 and Chk1 promoted 14-3-3 binding. We incubated purified Cdc25 with active or inactive versions of Cds1 and Chk1. Phosphorylated Cdc25 was then incubated with soluble Rad24, a fission yeast 14-3-3 protein, and interactions were monitored by immunoblotting. Phosphorylation of Cdc25 by either Chk1 or Cds1 promoted Rad24 binding *in vitro* (Fig. 3a, lanes 2, 4). The greater association between Rad24 and Cds1-phosphorylated Cdc25 reflects the greater stoichiometry of Cdc25 phosphorylation by Cds1, than by Chk1. Mutation of serines 99, 192 and 359 resulted in a threefold reduction in Rad24 binding, indicating that these sites are important for mediating interactions between Cdc25 and 14-3-3 proteins *in vitro* (Fig. 3b: compare lanes 1, 3 and 2, 4). Phosphorylated Cdc25 also bound Rad25, the other fission yeast 14-3-3 protein (data not shown).

To determine the role of phosphorylation on putative 14-3-3-binding sites in the checkpoint response to unreplicated DNA, we introduced plasmids encoding *cdc25+* and *cdc25-S3* into *cdc25-22*, a fission yeast strain containing a temperature-sensitive *cdc25+* allele. Both *cdc25+* and *cdc25-S3* fully rescued the cell-cycle arrest of *cdc25-22* cells at the restrictive temperature, and neither significantly advanced mitosis, as judged by the absence of small ('wee') cells in either culture (data not shown). To test the integrity of the checkpoint response, we incubated *cdc25+* and *cdc25-S3* transformants in hydroxyurea for 8 h and monitored the formation of

'cuts'. Fluorescence-activated cell sorting (FACS) analysis showed that the cells had a G1-phase DNA content under these conditions (data not shown), so any cuts observed were due to defects in the checkpoint response to unreplicated DNA. Mutation of canonical 14-3-3-binding sites measurably disrupted the checkpoint response (Fig. 4a): the *cdc25-S3* mutant showed substantially more 'cuts' (19% in the mutant compared to 2.1% in the wild-type). However, the checkpoint was not completely abolished, as many of the cells arrested normally. To determine whether the residual checkpoint response in *cdc25-S3* cells required either Mik1 or Wee1 (ref. 17), we studied the checkpoint response in *wee1⁻* and *mik1⁻* cells containing either *cdc25+* or the *cdc25-S3* allele. The formation of 'cuts' in *wee1⁻* *cdc25-S3* double mutants was not substantially stimulated by hydroxyurea (Fig. 4a), indicating that the checkpoint response of the *cdc25-S3* mutant is independent of Wee1. In contrast, combining the *cdc25-S3* mutation with loss of *mik1* caused a substantial checkpoint defect, with 37.8% of the cells forming 'cuts' after incubation in hydroxyurea (Fig. 4a, b). Thus, checkpoint arrest in response to unreplicated DNA requires negative regulation of Cdc25 by Cds1 and Chk1, as well as Mik1 function. Additional Cdc25 phosphorylation sites may contribute to the remaining checkpoint.

Our results establish that phosphorylation of Cdc25 at canonical 14-3-3-binding sites is an important component of the wild-type checkpoint response to unreplicated DNA. Replication-checkpoint defects in *rad24⁻* cells have been reported^{11,21}, consistent with our proposal that 14-3-3 proteins are required for the replication checkpoint. These results indicate that the DNA-replication and DNA-damage checkpoints use some of the same biochemical mechanisms to block entry into mitosis (Fig. 5). In response to unreplicated DNA, Cds1 or Chk1 phosphorylates Cdc25, creating 14-3-3-binding sites. The subsequent binding of 14-3-3 proteins prevents Cdc25 from activating Cdc2. Because either kinase can phosphorylate Cdc25, *cds1⁻* and *chk1⁻* single mutants do not have striking defects in the replication checkpoint. However, Cds1 also has a unique role in preventing DNA damage during S-phase arrest¹⁶. Levels of the Cdc2-inhibitory kinase Mik1 increase in a Cds1-dependent manner in hydroxyurea-treated cells¹⁷. The significant

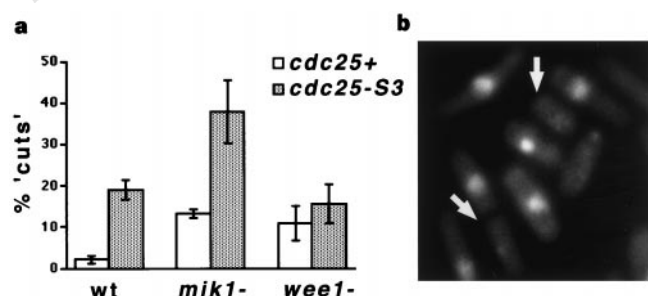


Figure 4 Mutation of canonical 14-3-3-binding sites of Cdc25 disrupts the checkpoint response to unreplicated DNA. **a**, Percentage of cuts in cultures of *cdc25-22* (strain TE282), *cdc25-22 wee1⁻* (TE899) or *cdc25-22 mik1⁻* (TE938) transformed with *cdc25+* (pTE525) or *cdc25-S3* (pTE547) after an 8-h incubation in 10 mM hydroxyurea. Four or five independent transformants were assayed for each data point; means and standard deviations are shown. **b**, Photograph of hydroxyurea-treated *cdc25-S3 mik1⁻* cells. 'Cuts' are indicated by arrows.

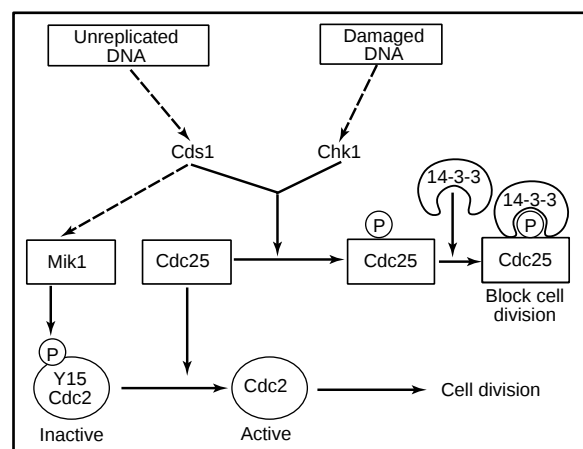


Figure 5 Targets of the replication checkpoint. The Cds1 and Chk1 kinases are effectors of the DNA-replication and -damage checkpoints, respectively. Cds1 and Chk1 functionally inactivate the phosphatase Cdc25 by promoting its binding to 14-3-3. The DNA-replication checkpoint is intact in cells lacking Cds1 because Chk1 functionally inactivates Cdc25 under these conditions. The kinase Mik1 inactivates the mitotic regulator Cdc2 by phosphorylating Cdc2 on tyrosine 15. Mik1 levels increase in a Cds1-dependent manner upon activation of the DNA-replication checkpoint¹⁷. Thus, the DNA-replication checkpoint blocks cell division by upregulating the Cdc2 inhibitor, Mik1 and by inactivating the Cdc2 activator, Cdc25.

checkpoint defect of strain *mik1⁻cdc25-S3* indicates that Mik1 and phosphorylation of Cdc25 function in parallel to ensure mitotic arrest in the presence of unreplicated DNA. □

Methods

Analysis of yeast strains. *cds1⁻* transformants were grown on minimal plates lacking thiamine for 48 h, replica-plated to minus-thiamine medium + 5 mM hydroxyurea, and photographed after 4 days of growth at 29 °C (Fig. 1c). *cdc25-22* transformants (Fig. 4) were grown at 32 °C for 40 h before adding 10 mM hydroxyurea. Under these conditions, cells transformed with vector alone arrest in G2 phase, forming highly elongated cells, before addition of hydroxyurea. We chose 32 °C because hydroxyurea-induced arrest is less efficient at 36.5 °C, the usual restrictive temperature for *cdc25-22*. For analysis of cuts, cells were grown in hydroxyurea for the indicated times, and then fixed and stained with 4,6-diamidino-2-phenylindole (DAPI) as described¹⁸. At least 100 cells were counted for each data point. Expression of the exogenous *cdc25⁺*, *chk1⁺* and *cds1⁺* genes was controlled by a weakened *nmt1⁺* promoter on REP81-based vector²². For details of strains and plasmids, see Supplementary Information.

Protein kinase assays. Kinase reactions consisted of bound or soluble forms of the kinases (~1 µg), 1–2 µg glutathione S-transferase (GST)–Cdc25 or mutant proteins bound to glutathione–agarose beads (Sigma), and 50 µl incomplete kinase buffer (40 mM Tris-HCl, pH 7.4, 10 mM MgCl₂) supplemented with 2 mM dithiothreitol (DTT), 10 µM ATP, and 30 µCi [γ -³²P]ATP (3,000 Ci mmol⁻¹; NEN). Reactions were incubated at 30 °C for 15–60 min followed by SDS–PAGE. Radiolabelled proteins were transferred to a nitrocellulose membrane and visualized by autoradiography.

Rad24-binding assays. JM109 cells were lysed in NETN buffer²³ supplemented with protease inhibitors and GST–Cdc25 was precipitated with glutathione–agarose beads at 4 °C for 1 h. Beads were washed twice with LiCl buffer²⁴ and once with incomplete kinase buffer (40 mM Tris-HCl, pH 7.4, 10 mM MgCl₂), and were then mixed with 0.5 ml incomplete kinase buffer supplemented with 2 mM DTT, 2 mM ATP and soluble GST–kinases (0.05–3.0 µg) at room temperature for 0.5–1 h. Beads were washed twice with LiCl buffer and once with binding buffer (50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 100 mM NaCl, 0.5% NP-40). Beads containing 0.5 µg GST–Cdc25 were incubated with 100 µl binding buffer containing 0.4–1 µg Rad24 at room temperature for 15–30 min and then washed three times with binding buffer. Proteins were resolved by SDS–PAGE and were visualized by western blotting using antibodies specific for GST or 14-3-3 (antibody K-19; Santa Cruz). Under these conditions, kinases did not bind Rad24, and Cds1-phosphorylated Cdc25 bound more than 25% of the input Rad24. Rad24 binding was quantified using a Molecular Dynamics fluorescence STORM 840 system.

Mapping studies. Tryptic phosphopeptides from labelled Cdc25 were subjected to two-dimensional phosphopeptide mapping and reverse-phase HPLC⁸ and two-dimensional phosphoamino-acid analysis²⁴. Selected HPLC fractions were digested with proline-specific endoproteinase⁸, or with 40 µg endoproteinase Glu-C (Boehringer) in 100 mM NH₄HCO₃ (pH 7.8) at 37 °C overnight. Selected HPLC fractions were subjected to manual Edman degradation⁸. For *in vivo* studies, Cdc25 was immunoprecipitated using a polyclonal antibody (BP-2).

Identification of phosphorylation sites. Phosphorylated serine 99 was eluted in HPLC fractions 73–75, depending on the experiment, and was present in both full-length Cdc25 and Cdc25(N306), indicating that the phosphorylation site was located within the first 306 amino acids of Cdc25. Edman degradation of the phosphopeptide resulted in release of radioactivity in the third cycle. Digestion of the tryptic phosphopeptide with either proline-specific endoproteinase or endoproteinase GluC resulted in an earlier elution of the phosphopeptide by HPLC, indicating the presence of both proline and glutamic acid residues. Substitution of alanine for serine at position 99 resulted in a loss of fractions 73–75.

Phosphorylated serine 359 was usually eluted in two HPLC fractions (between fractions 65 and 68) and was absent when phosphorylated Cdc25(N306) was analysed by HPLC, indicating that the phosphorylation site was located after amino acid 306. Edman degradation of the phosphopeptides present in the first and second fractions resulted in release of radioactivity in cycles 3 and 4. The sequence inclusive of and surrounding serine 359 is RRTQS₃₅₉ME. Because trypsin cleaves inefficiently when Arg or Lys residues

appear in tandem²⁵, phosphopeptides 3 and 5 probably arise because of partial proteolysis between R355 and R356. The HPLC elution profile did not change after digestion with either proline-specific endoproteinase or endoproteinase GluC. Substitution of alanine for serine at position 359 resulted in the loss of fractions 65–68.

Phosphorylated serine 192 was eluted in HPLC fraction 77 and was present in Cdc25(N306), indicating that the phosphorylation site was located within the first 306 amino acids of Cdc25. Edman degradation resulted in release of radioactivity in the first and third cycles, indicating either a single peptide phosphorylated on two sites or a mixture of phosphopeptides in this fraction. Substitution of alanine for serine at position 192 resulted in the loss of radioactivity released in cycle 1. For further details of the methods, see Supplementary Information and refs 26–30.

Received 15 June; accepted 14 August 1998.

- Hartwell, L. H. & Weinert, T. A. Checkpoints: controls that ensure the order of cell cycle events. *Science* **246**, 629–634 (1989).
- Enoch, T. & Nurse, P. Mutation of fission yeast cell cycle control genes abolishes dependence of mitosis on DNA replication. *Cell* **60**, 665–673 (1990).
- Jin, P., Gu, Y. & Morgan, D. O. Role of inhibitory Cdc2 phosphorylation in radiation-induced G2 arrest in human cells. *J. Cell Biol.* **134**, 963–970 (1996).
- O'Connell, M. J., Raleigh, J. M., Verkade, H. M. & Nurse, P. Chk1 is a wee1 kinase in the G2 DNA damage checkpoint inhibiting cdc2 by Y15 phosphorylation. *EMBO J.* **16**, 545–554 (1997).
- Rhind, N., Furnari, B. & Russell, P. Cdc2 tyrosine phosphorylation is required for the DNA damage checkpoint in fission yeast. *Genes Dev.* **11**, 504–511 (1997).
- Walworth, N., Davey, S. & Beach, D. Fission yeast chk1 protein kinase links the rad checkpoint pathway to cdc2. *Nature* **363**, 368–371 (1993).
- Walworth, N. C. & Bernards, R. rad-dependent responses of the chk1-encoded protein kinase at the DNA damage checkpoint. *Science* **271**, 353–356 (1996).
- Pe'g, C.-Y. et al. Mitotic- and G2-checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine 216. *Science* **277**, 1501–1505 (1997).
- Furnari, B., Rhind, N. & Russell, P. Cdc25 mitotic inducer targeted by Chk1 DNA damage checkpoint kinase. *Science* **277**, 1495–1497 (1997).
- Sanchez, Y. et al. Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation via Cdc25. *Science* **277**, 1497–1501 (1997).
- Al-Khodairy, F. et al. Identification and characterization of new elements involved in checkpoint and feedback controls in fission yeast. *Mol. Biol. Cell* **5**, 147–160 (1994).
- Murakami, H. & Okayama, H. A kinase from fission yeast responsible for blocking mitosis in S phase. *Nature* **374**, 817–819 (1995).
- Francesconi, S., Grenon, M., Bouvier, D. & Baldacci, G. p56^{chk1} protein kinase is required for the DNA replication checkpoint at 37 °C in fission yeast. *EMBO J.* **16**, 1332–1341 (1997).
- Uchiyama, M., Galli, I., Griffiths, D. J. F. & Wang, T. S.-F. A novel mutant allele of *Schizosaccharomyces pombe rad26* defective in monitoring S-phase progression to prevent premature mitosis. *Mol. Cell Biol.* **17**, 3103–3115 (1997).
- Sibon, O. C. M., Stevenson, V. A. & Theurkauf, W. E. DNA-replication checkpoint control at the *Drosophila* midblastula transition. *Nature* **388**, 93–97 (1997).
- Lindsay, H. D. et al. S phase-specific activation of Cds1 kinase defines a subpathway of the checkpoint response in *Schizosaccharomyces pombe*. *Genes Dev.* **12**, 382–395 (1998).
- Boddy, M. N., Furnari, B., Mondesert, O. & Russell, P. Replication checkpoint enforced by kinases Cds1 and Chk1. *Science* **280**, 909–912 (1998).
- Enoch, T., Carr, A. M. & Nurse, P. Fission yeast genes involved in coupling mitosis to completion of DNA replication. *Genes Dev.* **6**, 2035–2046 (1992).
- Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell* **84**, 889–897 (1996).
- Yaffe, M. B. et al. The structural basis for 14-3-3: phosphopeptide binding specificity. *Cell* **91**, 961–971 (1997).
- Forbes, K. C., Humphrey, T. & Enoch, T. Suppressors of Cdc25p overexpression identify two pathways that influence the G2/M checkpoint in fission yeast. *Genetics* (in the press).
- Basi, G., Schmid, E. & Maundrell, K. TATA box mutations in the *Schizosaccharomyces pombe nmt1* promoter affect transcription efficiency but not the transcription start point or the thiamine repressibility. *Gene* **123**, 131–136 (1993).
- Ogg, S., Gabrielli, B. & Piwnicka-Worms, H. Purification of a serine kinase that associates with and phosphorylates human Cdc25C on serine 216. *J. Biol. Chem.* **269**, 30461–30469 (1994).
- Liu, F., Stanton, J. J., Wu, Z. & Piwnicka-Worms, H. The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Mol. Cell Biol.* **17**, 571–583 (1997).
- van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by electrophoresis and chromatography on thin-layer cellulose plates. *Electrophoresis* **15**, 544–554 (1994).
- Moreno, S., Klar, A. & Nurse, P. In *Guide to Yeast Genetics and Molecular Biology* (eds Guthrie, C. & Fink, G. R.) 795–823 (Academic, San Diego, 1991).
- Lee, M. S., Enoch, T. & Piwnicka-Worms, H. Mik1 encodes a tyrosine kinase that phosphorylates p34cdc2 on tyrosine 15. *J. Biol. Chem.* **269**, 30530–30537 (1994).
- Maundrell, K. *nmt1* of fission yeast. *J. Biol. Chem.* **265**, 10857–10864 (1990).
- Kostrub, C. et al. Molecular analysis of hus1+, a fission yeast gene required for S-M and DNA damage checkpoints. *Mol. Gen. Genet.* **254**, 389–399 (1997).
- Moreno, S., Nurse, P. & Russell, P. Regulation of mitosis by cyclic accumulation of p80cdc25 mitotic inducer in fission yeast. *Nature* **344**, 549–552 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank H. Murakami and H. Okayama for pcL-cds1; A. Khokhlatchev and M. Cobb for pBB131; A. M. Carr for strains and plasmids; A. M. Carr and H. Lindsay for discussing construction of the *cds1⁻chk1⁻* mutant. This work was supported in part by the NIH. H.P.-W. is an Associate Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to H.P.-W. (e-mail: hpiwnica@cellbio.wustl.edu).

Efficiency of signalling through cytokine receptors depends critically on receptor orientation

Rashid S. Syed*, Scott W. Reid†, Cuiwei Li*, Janet C. Cheetham*, Kenneth H. Aoki*, Beishan Liu†, Hangjun Zhan†, Timothy D. Osslund*, Arthur J. Chirino*, Jiandong Zhang*, Janet Finer-Moore‡, Steven Elliott*, Karen Sitney*, Bradley A. Katz†, David J. Matthews†, John J. Wendoloski*, Joan Egrie* & Robert M. Stroud†‡

* Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320-1789, USA

† Axyx Pharmaceuticals Inc., 180 Kimball Way, South San Francisco, California 94080, USA

‡ Department of Biochemistry and Biophysics and Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143-0448, USA

Human erythropoietin is a haematopoietic cytokine required for the differentiation and proliferation of precursor cells into red blood cells¹. It activates cells by binding and orientating two cell-surface erythropoietin receptors (EPORs) which trigger an intracellular phosphorylation cascade². The half-maximal response in a cellular proliferation assay is evoked at an erythropoietin concentration of 10 pM (ref. 3), 10^{-2} of its K_d value for erythropoietin-EPOR binding site 1 ($K_d \approx 1$ nM), and 10^{-5} of the K_d for erythropoietin-EPOR binding site 2 ($K_d \approx 1$ μ M)⁴. Overall half-maximal binding (IC_{50}) of cell-surface receptors is produced with ~ 0.18 nM erythropoietin, indicating that only $\sim 6\%$ of the receptors would be bound in the presence of 10 pM erythropoietin. Other effective erythropoietin-mimetic ligands that dimerize receptors can evoke the same cellular responses^{5,6} but much less efficiently, requiring concentrations close to their K_d values (~ 0.1 μ M). The crystal structure of erythropoietin complexed

to the extracellular ligand-binding domains of the erythropoietin receptor, determined at 1.9 Å from two crystal forms, shows that erythropoietin imposes a unique 120° angular relationship and orientation that is responsible for optimal signalling through intracellular kinase pathways.

A soluble analogue of *Escherichia coli*-expressed erythropoietin (Swiss Prot accession no. P01588) was made by substituting lysine residues at three N-linked glycosylation sites (N24K, N38K and N83K). The extracellular ligand-binding domain of the erythropoietin receptor (EPObp; Swiss-Prot accession number P19235) was expressed in *Pichia pastoris* cells, with the N-linked glycosylation site removed by an N52Q mutation. An N164Q mutation was included to remove the potential for deamidation at the normal -N-G- sequence⁷. An A211E mutation was introduced for better expression and potentially enhanced folding efficiency on the basis of its role in improving the transportation of EPOR from the endoplasmic reticulum to the cell surface⁸.

Erythropoietin-(EPObp)₂ complex crystals diffracting to 2.8 Å ($a = 73.3$ Å, $b = 80.3$ Å, $c = 134.9$ Å, $P2_12_12_1$; Form 1) were obtained and the structure was determined by single isomorphous replacement (SIR) based on a mercury derivative, followed by solvent density modification by the methods of Abrahams⁹ in the SOLOMON algorithm (Table 1). This was alternated with density averaging of equivalent pairs of individual receptor domains. Higher-resolution data (1.9 Å) were obtained from a new crystal form, which used an erythropoietin analogue with two additional mutations, P121N and P122S, introduced into the CD loop. These changes were suggested on the basis of ¹⁵N NMR relaxation data, which showed extremely low-order parameters for the residues (Glu 117-Ala 128) in this loop with conformational heterogeneity in the backbone in the vicinity of the proline residues¹⁰. This analogue of erythropoietin, when complexed with EPObp, crystallized in a new unit cell ($a = 58.4$ Å, $b = 79.3$ Å, $c = 136.5$ Å, $P2_12_12_1$; Form 2) with a diffraction limit beyond 1.9 Å. The crystal structure of the erythropoietin-(EPObp)₂ complex (Form 2) was solved by multiple isomorphous replacement with anomalous scattering (MIRAS) using one uranium and two platinum derivatives (Table 1).

The structure of human erythropoietin complexed with EPObp

Table 1 Statistics for data collection and structure refinement.

Parameters	Native (Form 1)	HgAc ₂ (Form 1)	Native (Form 2)	UO ₂ (NO ₃) ₂ (Form 2)	K ₂ PtCl ₄ (Form 2)	K ₂ Pt(NO ₃) ₄ (Form 2)
Data collection						
Resolution (Å)	2.8	2.8	1.9	2.3	2.5	2.5
Total observations	88,794	105,221	251,681	252,332	167,497	137,452
Unique reflections	20,677	19,853	49,480	28,936	22,589	21,321
Average I/σ	8.6	9.3	19.1 (2.5)	12.0 (3.9)	12.7 (5.0)	12.4 (3.7)
Completeness (%)	91.6 (89.2)	88.0 (85.3)	97.0 (82.9)	99.7 (97.5)	99.9 (99.8)	93.4 (83.8)
R_{sym} (%) ^a	12.2 (37.4)	9.8 (38.2)	4.4 (37.7)	7.8 (46.9)	9.2 (46.9)	7.7 (44.8)
Phase determination						
Soaking conc. (mM)		1 and 10		0.1	0.1	4
Soaking time (h)		16		10	11	24
Number of sites		2		2	2	2
R_{iso} (%) ^b		7.3		7.4	12.7	12.0
R_c (%) ^c		64.3		65.0/43.0	62.0/44.0	65.0/44.0
Phasing power ^d		1.5/1.0		1.5/1.6	1.7/1.5	1.3/1.4
Refinement						
Resolution (Å)	80.0-2.8		50.0-1.9			
Reflections, $F > 2\sigma$	18,797		44,507			
$R_{\text{cryst}}/R_{\text{free}}^{\dagger}$ (%)	19.0/33.0		24.2/31.8			
No. water mols	0		300			
R.m.s.d. bond length (Å)	0.008		0.006			
R.m.s.d. bond angle (deg)	1.39		1.3			
B-factor (Å ²)	37.0		34.3			

Values given in parentheses are for highest-resolution shell. R.m.s.d., root-mean-square deviation.

^a $R_{\text{sym}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$, unweighted R -factor on I among symmetry mates.

^b $R_{\text{iso}} = \sum ||F_{\text{ph}}| - |F_{\text{p}}|| / \sum |F_{\text{ph}}|$, where F_{p} and F_{ph} are the structure factor amplitudes of the native and derivative data respectively.

^c $R_c = \sum ||F_{\text{ph}}| \pm |F_{\text{p}}|| - |F_{\text{h}}| / \sum ||F_{\text{ph}}| \pm |F_{\text{p}}||$ for centric reflections. The second entry is for anomalous dataset where $R_c = \sum ||F_{\text{ph}}^+| - |F_{\text{ph}}^-|| + ||F_{\text{ph}}^-| - |F_{\text{ph}}^+|| + |F_{\text{ph}}^-| + |F_{\text{ph}}^+|$ calculated for the top 25% largest Bijvoet differences. Data cutoff > 2.3 Å and $F/\sigma = 4.0$.

^d Phasing power = r.m.s. ($|F_{\text{h}}|/E$), where F_{h} is the calculated heavy-atom structure factor amplitude and E is the residual lack of closure error. The second entry is for anomalous phasing power, r.m.s. ($|F_{\text{h}}|/E_{\text{ano}}$), where F_{h} is the anomalous correction component amplitude. Data cutoff > 2.3 Å and $F/\sigma = 4.0$.

^e $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, where F is the structure factor.

^f $R_{\text{free}} = R$ factor computed for the test set of reflections (10% for Form 1 and 3% for Form 2) omitted from the refinement.

shows that one molecule of erythropoietin binds two receptors. In both crystal forms, the arrangement of molecules is similar in the *b/c* projection, although the packing of molecules is closer and different in the *a* direction in Form 2. The most significant difference is the positioning of the D2 domains relative to each other. In Form 1 there are no receptor–receptor contacts, whereas the more closely packed Form 2 structure has a hydrogen-bonding contact between the side chains of residues Ser 135 and Glu 134 on adjacent EPObp D2 domains. This is in sharp contrast to the analogous structure with human growth hormone (hGH) bound to its receptor (hGHbp), in which the D2 domains of the two receptor molecules interact extensively with a buried surface area of over 900 Å² (ref. 11). The Form 1 and Form 2 structures provide independent corroboration of key features of the complex. In particular, the overall structure from both crystal forms is very similar with respect to the interfaces between erythropoietin and EPObp. The angle between the adjacent EPObp D1 domains is the same, supporting the conclusion that the interfaces in the crystal structures represent those achieved by the complex at the cell surface.

Erythropoietin has an up–up–down–down four-helical bundle topology¹² with interhelical angles similar to those of the long-chain class, for example hGH and granulocyte colony-stimulating factor. However, it also contains two small antiparallel β -strands typical of the short-chain class¹³, for example macrophage colony-stimulating factor, stem-cell factor, interleukin-4 and interleukin-5. One pair of antiparallel long helices, α A (residues 8–26) and α D (residues 138–161), is held together by a disulphide bridge, Cys 7 to Cys 161. The other pair, α B (residues 55–83) and α C (residues 90–112), is linked by a short loop (Fig. 1). The α D helix is slightly irregular because of a small kink at Gly 151. The short segments of amino acids from the

long AB and CD crossover loops interact with each other to form an antiparallel β -sheet: β 1 (residues 39–41) and β 2 (residues 133–135). Several aromatic and hydrophobic residues, such as Phe 138, Phe 142, Tyr 145, Phe 148, Leu 153 and Tyr 156, on the interior face of the D-helix, pack against the non-polar side chains from the A, B and C helices to form the hydrophobic core of erythropoietin. A second disulphide bond, Cys 29 to Cys 33, links the end of the α A helix with part of the AB loop. Erythropoietin has two additional short helices, the α B' helix (residues 47–52) orthogonal to α B and the mini-helix α C' (residues 114–121) following α C with a 90° tilt beginning at Gly 113.

The structure of EPObp is composed of a short amino-terminal helix (~15 residues) followed by two β -sandwich domains, D1 (N-terminal) and D2 (carboxy-terminal), each consisting of ~100 residues. The overall fold of EPObp is as described previously for the crystal structure of the erythropoietin receptor (EBP) complexed to a peptide agonist (EMP1)⁶. The N-terminal helix (residues 9–22) of both EPObp structures is tucked into the elbow formed by the D1 and D2 domains (Figs 1 and 2a), in close proximity to the WSXWS motif (residues 209–213; X = Glu) in D2. The side chain of helix residue Leu 18 makes hydrophobic contacts (<4.5 Å) with Phe 29, Leu 27, Leu 120 and Phe 208 of EPObp. Alteration of the WSXWS sequence disrupts erythropoietin binding and receptor signalling¹⁴, correlating with similar findings for the interleukin-2, prolactin and hGH receptors¹⁵. The WSXWS motif has been shown to be critical for the folding and transport of the receptor to the cell surface, and an A211E mutation further improved the efficiency of the processes⁸. In the erythropoietin–(EPObp)₂ complex, the side chain of Glu 211 from the WSXWS motif is closest (<4.5 Å) to Leu 17 of the N-terminal receptor helix. Additionally, the side chains of residues Trp 209 and Trp 212 sandwich the hydrophobic side chain of Arg 197 in the receptor fold, whereas Ser 210 and Val 196, respectively. On the basis of the interactions of the WSXWS box with the N-terminal helix and the β -sheet residues Val 196, Arg 197 and Ala 198 in the erythropoietin–(EPObp)₂ complex, we suggest that the N-terminal helix is biologically important in stabilizing the folded EPOR. This helix position differs from the N-terminal helices seen in the two-fold symmetric EMP1–EBP complex, in which the connection between helix and D1 was not visible and unconnected helices were farther away from the D1/D2 elbow (Fig. 2b)⁶.

In the erythropoietin–(EPObp)₂ complex, the receptor molecules are held together through two regions located on opposite faces of erythropoietin (Figs 1 and 2a and Table 2). These erythropoietin–EPObp interfaces have been identified as high-affinity ($K_d \approx 1$ nM) and low affinity ($K_d \approx 1$ μ M) and are referred to as site 1 and site 2 respectively⁴. Site 1 is composed of erythropoietin residues from helices α A, α B', α D and part of the AB loop. Site 2 erythropoietin residues are located in α A and α C. Complementary receptor interfaces involve contributions from six EPObp loops (L1–L6) between β -strands from both domains: L1, L2 and L3 (D1 domain); L4 from the polypeptide D1–D2 linker; and L5 and L6 (D2 domain). Residues from all six loops are involved in the site 1 interface and all except the fourth loop participate at site 2.

Site 1 is characterized by a central hydrophobic binding pocket, flanked at opposite ends by hydrophilic interactions (Table 2). EPObp loops L1, L5 and L6 accommodate the hydrophobic erythropoietin residues, with Phe 93 of EPObp dominating the non-polar interactions. This pivotal residue is firmly held in place by hydrogen bonding to erythropoietin residues Thr 44 and Asn 147. The plane of the phenyl ring of Phe 93 packs against the C α of Gly 151 (at a distance of 3.8 Å) and is accommodated at a hydrophobic surface created by a quartet of erythropoietin side chains (Gly 151, Val 46, Phe 48 and Leu 155). Mutagenesis has shown that Phe 93 on the receptor is a critical determinant of erythropoietin binding; the F93A mutation eliminates any detectable binding^{16,17}. The site 1

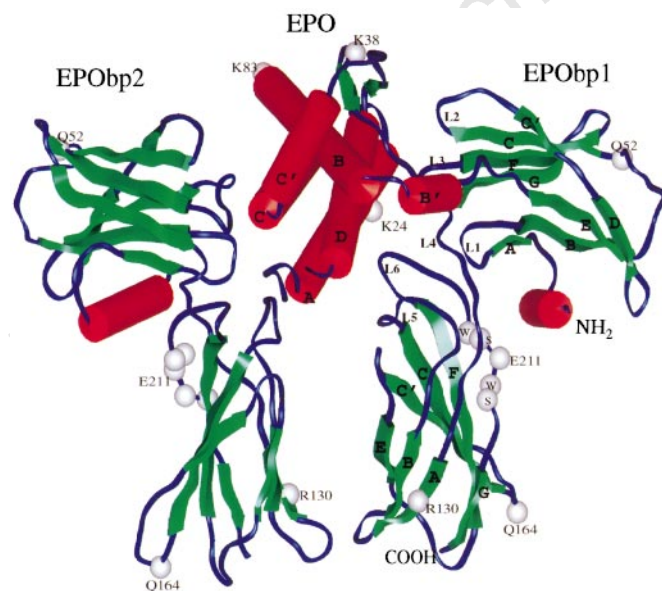


Figure 1 Crystal structure of the erythropoietin–(EPObp)₂ complex. α -Helices are shown as red cylinders and β -sheets as green ribbons. Erythropoietin: α A(8–26) up, α B'(47–52), α B(55–83) up, α C(90–112) down, α C'(114–121), α D(138–161) down, β 1(39–41), β 2(133–135). Disulphide bridges: Cys 7–Cys 161, Cys 29–Cys 33. D1 domain of EPObp: α 1(9–22), β A(26–30), β B(36–42), β C(52–59), β C'(65–68), β D(69–74), β E(78–85), β F(95–103), β G(106–113). Disulphide bridges: Cys 28–Cys 38, Cys 67–Cys 83. D2 domains of EPObp: β A(119–132), β B(137–143), β C(154–162), β C'(170–174), β E(179–183), β F(190–201), β G(216–220). EPObp loops that interact with erythropoietin are labelled L1(31–35), L2(60–64), L3(86–94), L4(114–118), L5(144–153) and L6(202–205). The locations of residues in the WSXWS box, dimerizing mutation site²³ R130C, two Asn $\rightarrow$ Gln mutation sites (52, 164) on EPObp, and three Asn $\rightarrow$ Lys mutations (24, 38, 83) on erythropoietin are depicted as white spheres. EPO, erythropoietin.

interface also has contributions from the A–B linker polypeptide residues (Thr 44–Phe 48) of erythropoietin. A smaller contribution is made by Met 150 of EPObp, which is within van der Waals contact distance of only a single amino acid, Phe 48, of erythropoietin. The hydrophobic binding surface (site 1) itself is surrounded by hydrophilic interactions (Table 2). Asn 147 of erythropoietin is central to the hydrophilic D helix interactions. It has three hydrogen bonds to EPObp residues, one of which is between its N δ atom and the carbonyl oxygen of Phe 93. On either side of the Asn 147 are arginine residues, Arg 143 and Arg 150, that interact with glutamic acids Glu 60 (L2) and Glu 117 (L4), respectively. Flanking the binding site at one end, Ser 9 and Arg 10 from the A helix of erythropoietin interact with Glu 176 and His 153, located at the end of β C' and in the L5 loop of the receptor, respectively. At the opposite end, Arg 143 and Lys 140 of α D, and Lys 45 of the AB loop of erythropoietin, interact with residues Glu 60, Asp 61 and Glu 62 of the L2 loop of the EPObp D1 domain.

At site 2, the interactions are less extensive than at site 1. Most of the interactions are between residues of the C helix of erythropoietin and the L3 loop of EPObp. The hydrophobic surface is created on EPObp by residues Phe 93, Phe 205 and Met 150, although because of the positioning of the D1 domain relative to the interaction of site 2 with erythropoietin there is a relatively flat surface that erythropoietin residues can interact with. The closest non-polar contact is between erythropoietin Leu 108 and EPObp Phe 93 C γ (3.9 Å). The rest of the C helix residues are positioned at a greater distance from the EPObp hydrophobic surface than at the site 1 interface. Phe 93 is again the major contributor to the central hydrophobic binding pocket (Leu 5, Val 11, Tyr 15, Ser 104, Thr 107 and Leu 108 of erythropoietin are within 4.5 Å of EPObp Phe 93). Met 150 of EPObp, however, is more buried than in site 1, making van der Waals contacts with Arg 10, Val 11 and Arg 14 of erythropoietin. A total of 12 hydrophilic interactions are formed between residues in the A and C helices of erythropoietin (Asp 8, Arg 14, Lys 97, Ser 100,

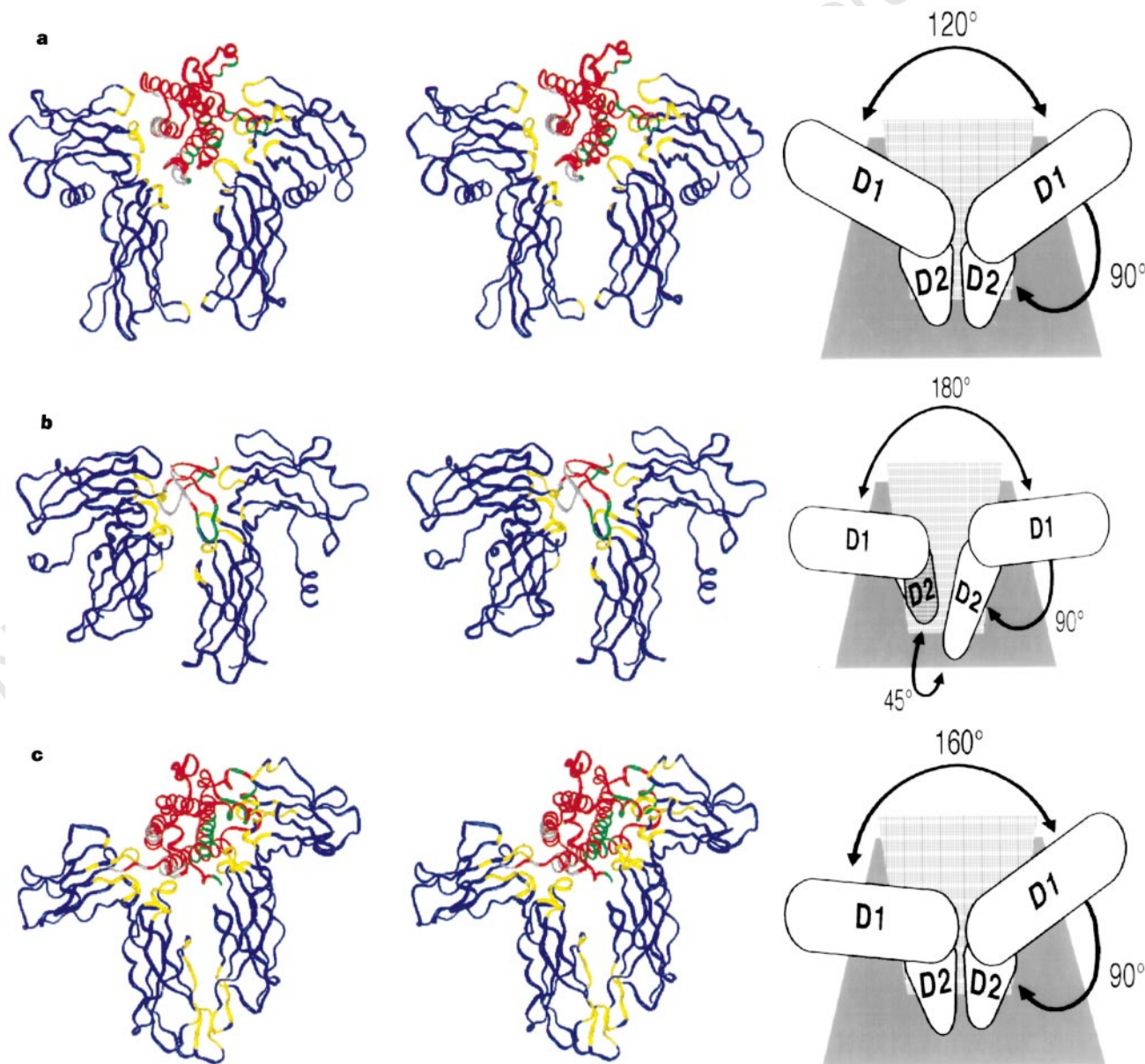


Figure 2 Stereo views. **a**, erythropoietin–(EPObp)₂. **b**, (EMP1–EBP)₂. **c**, hGH–(hGHbp)₂. All the receptors are coloured blue, with their ligands in red. The intermolecular contacts (<4.5 Å) involving receptor amino acids are shown in yellow; ligand amino acids with the receptor on the right-hand side (site 1 for erythropoietin and hGH) are shown in green; and ligand amino acids with the

receptor on the left-hand side (site 2 for erythropoietin and hGH) are shown in white. Receptor orientations are also represented schematically at the right, with the horizontal membrane plane shown in grey and the corresponding vertical plane drawn in a chequered pattern.

Table 2 Hydrogen bonds and salt bridges between erythropoietin (EPO) and EPObp in sites 1 and 2 of the EPO–(EPObp)₂ complex

Site 1 atoms			Site 2 atoms		
EPO	EPObp	Distance (Å)	EPO	EPObp	Distance (Å)
Lys 20 N ζ	Glu 202 O ϵ 1	3.5	Asp 8 O δ 2	His 153 Ne2	2.9
Thr 44 O γ	Phe 93 O	2.6	Arg 14 N η 2	Leu 33 O	3.1
Lys 45 N ζ	Glu 62 O ϵ 1	3.1	Lys 97 N ζ	Glu 34 O ϵ 1	2.7
Val 46 N	Ser 92 O	3.0	Ser 100 O	Ser 91 O γ	2.9
Val 46 O	Ser 92 N	2.8	Arg 103 Ne	Glu 62 O ϵ 1	3.1
Asn 47 N δ 2	Thr 87 O	2.9	Arg 103 Ne	Glu 62 O ϵ 2	3.5
Asn 47 N δ 2	Ala 88 O	3.2	Arg 103 N η 1	Ser 91 O γ	2.7
Arg 131 N η 2	Asp 61 O	2.8	Arg 103 N η 1	Ala 88 O	2.6
Lys 140 N ζ	Asp 61 O δ 2	3.0	Arg 103 N η 2	Ala 88 O	3.3
Arg 143 N η 2	Glu 60 O ϵ 2	2.8	Arg 103 N η 2	Asp 89 O δ 1	3.1
Asn 147 N δ 2	Phe 93 O	3.1	Ser 104 O γ	Ser 92 O	2.9
Asn 147 N δ 2	His 114 Ne2	3.0			
Asn 147 O δ 1	His 114 Ne2	3.5			
Arg 150 O	Ser 204 O γ	2.6			
Arg 150 N η 1	Pro 203 O	2.8			
Arg 150 N η 1	Glu 117 O ϵ 2	3.2			
Arg 150 N η 2	Glu 117 O ϵ 2	2.8			

Arg 103 and Ser 104) and loops L1, L2, L3 and L5 of EPObp (Leu 33, Glu 34, Glu 62, Ala 88, Asp 89, Ser 91, Ser 92 and His 153), with a total of 11 hydrogen bonds (Table 2). The side chains of helix A residues Asp 8 and Arg 14 form hydrogen bonds with the side chain of His 153 and the carbonyl oxygen of Leu 33, respectively.

In general, the sidechain interactions at both interfaces are predominantly between positively charged lysine and arginine residues of erythropoietin and negatively charged aspartate and glutamate side chains of EPObp. Site 1 contains almost twice as many sidechain–sidechain interactions as site 2. However, the numbers of hydrophilic contacts involving main-chain atoms from either erythropoietin or EPObp are equal for both sites. In addition, site 1 contains two hydrogen bonds that involve main-chain atoms from both erythropoietin and EPObp. Hydrophilic interactions from three EPObp residues (Glu 62, Ala 88 and Ser 92) are common to both interfaces, two of which involve hydrogen bonding from a main-chain atom (Table 2). Although most residues that interact across both interfaces are polar (Table 3), the expected contribution to binding affinity is greatest from the hydrophobic contacts.

Although the crystal structure of erythropoietin–(EPObp)₂ clearly shows an extensive interface between erythropoietin and its receptor (buried surface areas at sites 1 and 2 are 920 and 660 Å²,

respectively), a much smaller subset of residues, identified by mutagenesis, delineates a functional epitope (Table 3)^{18–21}. Gly 151 has an important structural role in erythropoietin, forming a kink in the D helix that brings the side chain of Lys 152 into hydrophobic contact with Val 63, Trp 51 and Phe 148 in the protein core. Replacement of alanine at either position 151 or 152 causes a substantial loss of bioactivity^{19–21}. The basic residues Lys 20 and Lys 45 on erythropoietin in the interface of site 1 are unaffected by Ala substitutions¹⁹, but mutations to acidic residues do cause substantial loss of bioactivity²¹. Arg 103 of erythropoietin has been implicated in site 2 binding^{18–21} and from the structure we see that it is completely buried at the interface, interacting with six residues on EPObp. Erythropoietin mutant R103A has been shown to have unchanged site 1 affinity, but to lack proliferative function owing to loss of site 2 binding²⁰. A sequential binding mechanism was proposed on the basis of this analysis of several erythropoietin mutants. Arg 14 on erythropoietin, which interacts with four residues on EPObp, is also very sensitive to mutagenesis^{19–21}. Sedimentation equilibrium data have shown that the R14Q mutation in erythropoietin reduces the site 2 affinity roughly fivefold (J. Philo, personal communication).

Comparison of the erythropoietin–(EPObp)₂ interactions with those of hGH–(hGHbp)₂ (ref. 11) reveals a number of significant

Table 3 List of the buried residues (>5.0 Å²) (one-letter codes) of erythropoietin EPO in site 1 and site 2 intermolecular contact areas

Site 1				Site 2			
EPO residue	Buried surface area (Å ²)*	Neighbouring EPObp residues (4.5-Å cutoff)	Effect of EPO mutagenesis on bioactivity†	EPO residue	Buried surface area (Å ²)*	Neighbouring EPObp residues (4.5-Å cutoff)	Effect of EPO mutagenesis on bioactivity†
S9	17	H153	A, N	L5	55	F93, S204	
R10	24	H153, E176	A, <u>I</u>	D8	23	H153	S
E13	30	P203	A	R10	51	M150, S152, H153	A, <u>I</u>
L16	22	P203, S204	A, S	V11	49	F93, M150, H153, S204	S
L17	15	P203	A, S	R14	108	L33, E34, P149, M150	L, A, E, Q
K20	48	E202, P203	A, I, R, E	Y15	16	S92, F93	F, <u>I</u>
T44	19	S92, F93, V94	I	Q78	11	A88	A, E, R
K45	107	E62, S91, S92, F93, V94	A, R, I, D	D96	44	T87, A88	A, <u>R</u>
V46	38	T90, S91, S92, F93	A	K97	36	E34	R, A, E
N47	67	T87, A88, D89, T90, S91	A	V99	7	A88	A, S
F48	119	L33, E34, S92, F93, F205	I, S, G	S100	40	T87, A88, T90, S91	A, <u>R, E, T</u>
Y49	61	E34, T87	F, S	R103	118	L59, E62, A88, D89, S91, V94	K, A, E, H, N, Q
K52	17	E34	S, R, Q, <u>E</u>	S104	38	S91, S92, F93	A, I
R131	34	E60, D61, E62	T	T107	50	F93, V94	A, L, S
I133	22	D61		L108	15	F93	A, K
K140	31	E60, D61	A, R, M, T	R110	26	E60, D61	T, E
R143	43	E60, P95	E, A				
N147	43	F93, V94, P95, H114	<u>A, K</u>				
R150	128	F93, H114, N116, E117, P203, S204	A, Q, E				
G151	14	F93	A				
K154	32	H153, S204	A, R, S				
L155	26	F93	A, N				

* The solvent-accessible surface areas were calculated by using XPLO3.851 with 1.60-Å probe sphere and standard atomic radii.

† Available mutagenesis data^{18–21} shown with substituted amino acids highlighted according to degree of loss of *in vitro* bioactivity: bold underlined, >50 times; bold, >5 times; underlined, 2–5 times; unhighlighted, no effect.

differences. In hGH-(hGHbp)₂, two different regions of the AB loop of hGH (the start of the loop at Gln 46 and the end of the loop at Arg 64) are involved in site-1 ligand-receptor interactions. In contrast, only the C-terminal end of the AB loop of erythropoietin, at residue Arg 45, is involved in site-1 erythropoietin-EPObp binding. In the hGH-(hGHbp)₂ complex, two hGHbp residues, Trp 104 and Trp 169, form a central hydrophobic pocket flanked by hydrophilic interactions. In comparison, in the erythropoietin-(EPObp)₂ complex, Phe 93, the direct homologue of Trp 104 in hGHbp, is the primary contributor to the interactions between receptor and ligand in the hydrophobic pocket; Met 150 on EPObp (homologous to Trp 169 on hGHbp) is more peripherally involved. As a result of these differences, the binding between ligand and receptor mediated by the central hydrophobic pocket of the receptor seems to be more localized for erythropoietin-(EPObp)₂ than for hGH-(hGHbp)₂. This region has been identified as a mutational hot spot in hGH-(hGHbp)₂ (ref. 22).

Comparison of the erythropoietin-(EPObp)₂, (EMP1-EBP)₂ and hGH-(hGHbp)₂ structures reveals that the receptor molecules have significantly different orientations relative to each other (Fig. 2a-c). Viewed perpendicular to the membrane plane, the D1 domains from the two receptor molecules in the erythropoietin-(EPObp)₂ complex are positioned ~120° relative to each other, compared with 180° in (EMP1-EBP)₂. For the hGH-(hGHbp)₂ complexes, the angle is 160°. The two D2 domains in the erythropoietin-(EPObp)₂ and hGH-(hGHbp)₂ complexes are in approximately the same plane, perpendicular to the membrane. In contrast, the two D2 domains in the (EMP1-EBP)₂ complex are twisted away from each other by ~45°.

Disulphide crosslinking between murine erythropoietin receptors, generated by the replacement of Arg 130 with cysteine, resulted in constitutive activation²³. The two EPObp Arg 130 residues, located near the membrane-proximal region of the D2 domains in the two adjacent receptors, are 27 Å apart. Formation of a disulphide bond would therefore require significant angular rearrangement of the receptor domains. Signalling in this case might be the result of a subset of the receptor molecules dimerizing intracellularly before being transported to the cell surface, as only a fraction of EPOR achieves constitutive activation²³. Alternatively, a constrained angular relationship in the mutant murine dimers could bring about an impaired signalling event while still allowing erythropoietin to bind to its receptors.

The efficiency of signalling by erythropoietin mimetics is typically at most 1/20 that of the natural ligand³. Our results therefore show that the activating efficiency of EPOR, and by inference each of the cytokine receptor complexes, depends critically on the separation, orientation and relative disposition of bound receptors, and hence the effect on their unique intracellular domains. This consideration is therefore an important parameter in the design of erythropoietin mimetics; specifically, it suggests that asymmetrical molecules might be required to achieve these orientations most efficiently. □

Methods

Protein preparation and crystallization. Recombinant human erythropoietin mutants, r-met lys hu erythropoietin N24K N38K N83K and r-met lys hu erythropoietin N24K N38K N83K P121N P122S, were produced in *E. coli* cells and purified as described²⁴. Recombinant human EPObp (r-REF huEPOR N52Q N164Q A211E) was secreted from *Pichia pastoris*, and purified as described (H.Z. *et al.*, manuscript in preparation). EPObp contains three N-terminal residues (Arg-Glu-Phe) that replace Ala in the wild-type sequence²⁵. Both erythropoietin-(EPObp)₂ complexes were crystallized by hanging-drop vapour diffusion at 20 °C. Protein was mixed with an equal volume of the crystallization-well buffer containing 15% PEG 4K, 0.2 M CaCl₂ and 0.1 M MES pH 6.5–7.0 (Form 1) and in cryoprotectant conditions of 32% PEG 1500, 0.28 M (NH₄)₂SO₄, 0.1 M MES, pH 6.5 (Form 2). Form-1 data from three isomorphous crystals (collected at 20 °C) constituted the initial native data set. Data were collected with an R-AXIS II image plate system mounted on a Rigaku

RU200 generator using a rotating copper anode target tube operating at 50 kV and 60 mA. These were autoindexed, integrated, scaled and merged with the R-AXIS associated data-reduction software BIOTEX (Molecular Structure Corporation). Form-2 data were collected with a single crystal at 100K on an R-AXIS IV imaging system installed on an 18 kW Rigaku RU300 generator. Data were indexed, integrated and scaled with the programs DENZO and SCALEPACK²⁶.

Structure determination and refinement. For Form 1, two highly anisotropic mercury sites from the single derivative (Table 1) were refined with MLPHARE²⁷ and double-difference maps (mean figure of merit = 0.31). Density modification, with SOLOMON⁹ and DM²⁷, yielded a map at 4 Å resolution that outlined the α-helices in erythropoietin and the pattern of β-sheets in the ordered chains of EPObp. 300 cycles of density modification followed as the resolution was extended from 4.0 to 2.8 Å in 0.004-Å steps. At each step the solvent mask was recalculated by using a radius equal to the resolution of the map. Structure factors calculated from the modified map were used to estimate the phases in the next resolution shell. These were not combined with the SIR phase probabilities owing to the low phasing power beyond 4 Å. Further to resolution extension, the density for domains D1 and D1', D2 and D2' were averaged to improve the phases iteratively by imposing constraints of equivalence in the domains. The D1 and D2 domain boundaries for each receptor were identified by inspection and the rotation angles between similar domains were determined by maximizing the correlation between the selected map regions. The initial structure solution gave an *R*_{free} of 46%. The structure was refined with XPLO3.851 (ref. 28) interleaved with density averaging between the D1/D1' and D2/D2' domains and calculation of omit maps. Several rounds of refinement and building to 'complete' omit and 2*F*_o - *F*_c maps resulted in the final model (*R*_{cryst}/*R*_{free} = 19.0%/33.0%). Towards the latter stages of refinement, the release of the (EMP1-EBP)₂ coordinates (Brookhaven Protein Data Bank (PDB) accession number 1EBP) allowed the positioning of some surface exposed loops in the EPObp receptor.

For Form 2, heavy-atom sites were identified for each derivative in difference Patterson maps, refined for their positions and occupancies, and confirmed by examining the peaks in cross-phased difference Fourier maps²⁹. The MIRAS phases calculated²⁹ by using isomorphous and anomalous differences from the three derivatives (Table 1) with the native data (mean figure of merit 0.72, 50.0–2.3 Å) and solvent flattened with SIGMAA weighting (mean figure of merit 0.92, 50.0–2.3 Å) produced a good-quality map. This map was used to build the complete model of erythropoietin-(EPObp)₂ with the program package QUANTA97 (Molecular Simulations). The starting templates for the D1 and D2 domains of EPObp used in model building were taken from the preliminary crystal structure of the (EMP1-EBP)₂ complex (T.D.O. *et al.*, manuscript in preparation) reproduced from previously published work⁶. The locations of cysteine residues were confirmed by examining an anomalous map, which showed clear peaks for sulphur atoms contoured at 2.0σ. The structure was refined with 705 reflections (3%) excluded for cross-validation (*R*_{free}) by using simulated annealing accompanied by the bulk solvent correction protocol in XPLO3.851 (ref. 28). The model after nine cycles of manual model building, simulated annealing, and positional and temperature factor refinements yielded an *R*_{cryst} and an *R*_{free} of 0.236 and 0.339 (50.0–2.3 Å, *F* > 2σ_i), respectively. An additional three cycles of model building, refinement and phase combination at 1.9 Å resolution with 1323 reflections (3%) excluded for cross-validation (*R*_{free}) resulted in a crystallographic *R*_{cryst}/*R*_{free} of 0.242/0.318 (50.0–1.9 Å, *F* > 2σ_i).

Received 20 May; accepted 11 August 1998.

- Graber, S. E. & Krantz, S. B. Erythropoietin and the control of red blood cell production. *Annu. Rev. Med.* **29**, 51–66 (1978).
- Damen, J. E. & Krystal, G. Early events in erythropoietin-induced signaling. *Exp. Hematol.* **24**, 1455–1459 (1996).
- Johnson, D. L. *et al.* Identification of a 13 amino acid peptide mimetic of erythropoietin and description of amino acids critical for the mimetic activity of EMP1. *Biochemistry* **37**, 3699–3710 (1998).
- Philo, J. S., Aoki, K. H., Arakawa, T., Narhi, L. O. & Wen, J. Dimerization of the extracellular domain of the erythropoietin (EPO) receptor by EPO: one high-affinity and one low-affinity interaction. *Biochemistry* **35**, 1681–1691 (1996).
- Wrighton, N. C. *et al.* Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458–464 (1996).
- Livnah, O. *et al.* Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464–471 (1996).
- Derby, P., Aoki, K. H., Katta, V. & Rohde, M. F. in *Techniques in Protein Chemistry VII* (ed. Marshak, D. R.) 109–119 (Academic, San Diego, 1996).
- Hilton, D. J., Watowich, S. S., Katz, L. & Lodish, H. F. Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor. *J. Biol. Chem.* **271**, 4699–4708 (1996).

9. Abrahams, J. P. & Leslie, A. G. W. *Acta Crystallogr. D* **52**, 30–42 (1996).
10. Cheetham, J. C. *et al.* NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nature Struct. Biol.* (in the press).
11. De Vos, A. M., Ultsch, M. & Kossiakoff, A. A. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255**, 306–312 (1992).
12. Sprang, S. R. & Bazan, J. F. Cytokine structural taxonomy and mechanisms of receptor engagement. *Curr. Opin. Struct. Biol.* **3**, 815–827 (1993).
13. Rozwarski, D. A. *et al.* Structural comparisons among the short-chain helical cytokines. *Structure* **2**, 159–173 (1994).
14. Yoshimura, A. *et al.* Mutations in the Trp-Ser-X-Trp-Ser motif of the erythropoietin receptor abolish processing, ligand binding, and activation of the receptor. *J. Biol. Chem.* **267**, 11619–11625 (1992).
15. Baumgartner, J. W., Wells, C. A., Chen, C. M. & Waters, M. J. The role of the WSXWS equivalent motif in growth-hormone receptor function. *J. Biol. Chem.* **269**, 29094–29101 (1994).
16. Barbone, F. P. *et al.* Mutagenesis studies of the human erythropoietin receptor. Establishment of structure–function relationships. *J. Biol. Chem.* **272**, 4985–4992 (1997).
17. Middleton, S. A. *et al.* Identification of a critical ligand-binding determinant of the human erythropoietin receptor—evidence for common ligand-binding motifs in the cytokine receptor family. *J. Biol. Chem.* **271**, 14045–14054 (1996).
18. Grodberg, J., Davis, K. L. & Skowski, A. J. Alanine scanning mutagenesis of human erythropoietin identifies four amino acids which are critical for biological activity. *Eur. J. Biochem.* **218**, 597–601 (1993).
19. Wen, D. Y., Boissel, J. P., Showers, M., Ruch, B. C. & Bunn, H. F. Erythropoietin structure–function relationships—identification of functionally important domains. *J. Biol. Chem.* **269**, 22839–22846 (1994).
20. Matthews, D. J., Topping, R. S., Cass, R. T. & Giebel, L. B. A sequential dimerization mechanism for erythropoietin receptor activation. *Proc. Natl Acad. Sci. USA* **93**, 9471–9476 (1996).
21. Elliott, S., Lorenzini, T., Chang, D., Barzilay, J. & Delorme, E. Mapping of the active site of recombinant human erythropoietin. *Blood* **89**, 493–502 (1997).
22. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone–receptor interface. *Science* **267**, 383–386 (1995).
23. Watowich, S. S. *et al.* Homodimerization and constitutive activation of the erythropoietin receptor. *Proc. Natl Acad. Sci. USA* **89**, 2140–2144 (1992).
24. Narhi, L. O. *et al.* The effect of carbohydrate on the structure and stability of erythropoietin. *J. Biol. Chem.* **266**, 23022–23026 (1991).
25. Johnson, D. L. *et al.* Refolding, purification, and characterization of human erythropoietin binding protein produced in *Escherichia coli*. *Prot. Express. Purif.* **7**, 104–113 (1996).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
28. Brunger, A. T. *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).
29. Furey, W. & Swaminathan, S. Phases-95: a program package for processing and analyzing diffraction data from macromolecules. *Methods Enzymol.* **277**, 590–620 (1997).

Acknowledgements. We thank J. Philo, M. McGrath, P. Sprengler, W. Welch, J. Rupert, L. Narhi, G. Rogers, M. Rohde, S. Jordan, K. Langley, R. Mackman and M. Venuti for valuable discussions. Structure analyses of Form 1 and Form 2 were independently performed at Axyx Pharmaceuticals Inc. and Amgen Inc., respectively.

Correspondence and requests for materials should be addressed to R.S.S. (e-mail: rsyed@amgen.com) or R.M.S. (e-mail: stroud@msg.ucsf.edu). Coordinates have been deposited with the Brookhaven Protein Data Bank (accession numbers 1blw for Form 1 and 1eer for Form 2).

Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins

Kerstin Hill*, Kirstin Model†, Michael T. Ryan†, Klaus Dietmeier†, Falk Martin†‡, Richard Wagner* & Nikolaus Pfanner†

* Biophysik, Universität Osnabrück, FB Biologie/Chemie, D-49034 Osnabrück, Germany

† Institut für Biochemie und Molekularbiologie, Universität Freiburg, Hermann-Herder-Strasse 7, D-79104 Freiburg, Germany

‡ Fakultät für Biologie, Universität Freiburg, D-79104 Freiburg, Germany

The mitochondrial outer membrane contains machinery for the import of preproteins encoded by nuclear genes^{1–3}. Eight different Tom (translocase of outer membrane) proteins have been identified that function as receptors and/or are related to a hypothetical general import pore. Many mitochondrial membrane channel activities have been described^{4–7}, including one related to Tim23 of the inner-membrane protein-import system⁵; however, the pore-forming subunit(s) of the Tom machinery have not been identified until now. Here we describe the expression and functional reconstitution of Tom40, an integral membrane protein with mainly β -sheet structure. Tom40 forms a cation-selective high-conductance channel that specifically binds to and trans-

ports mitochondrial-targeting sequences added to the *cis* side of the membrane. We conclude that Tom40 is the pore-forming subunit of the mitochondrial general import pore and that it constitutes a hydrophilic, ~ 22 Å wide channel for the import of preproteins.

We expressed *Saccharomyces cerevisiae* Tom40 (ref. 8) in *Escherichia coli* cells (Fig. 1a, lane 2). Tom40, representing $\sim 25\%$ of total *E. coli* protein, accumulated in inclusion bodies and could be solubilized with urea. As Tom40 may be slightly similar to porins at the secondary structure level⁹, we modified a method for the preparation and renaturation of *Rhodospseudomonas blastica* porin from *E. coli* inclusion bodies¹⁰. Tom40 was isolated to high purity ($>95\%$) (Fig. 1a, lane 3). The two protein bands of minor abundance that are visible in the purified preparation of Tom40 (Fig. 1a, lane 3) are degradation products of Tom40. When diluted into the non-ionic detergent nonanoyl-*N*-methylglucamide (Mega-9), the recombinant Tom40 migrated as a single band on blue native electrophoresis and this band was indistinguishable from that of Tom40 that had been solubilized from mitochondria by Mega-9 (Fig. 1b). Circular dichroism analysis of Tom40 in urea and after dilution into Mega-9 indicated an efficient refolding of the protein upon dilution (Fig. 1c). We inserted Tom40 into liposomes by dilution of the urea-denatured protein into a mixture of Mega-9 and azolectin, with subsequent removal of the detergent. The circular dichroism spectrum of liposome-inserted Tom40 was comparable to that of Tom40 in Mega-9 (Fig. 1c). A secondary structure calculation according to ref. 11 indicated a predominance of β -sheet structure ($>60\%$) and little α -helical structure ($<5\%$) in the protein.

To assess whether Tom40 was properly reconstituted in the liposomes, we determined its topological orientation in liposomes in comparison to that of Tom40 in mitochondria. Treatment of mitochondria with trypsin cleaved a small peptide (relative molecular mass (M_r) ~ 2.5 K) from Tom40 (Fig. 1d, upper panel, lanes 2, 3). The same cleavage occurred after recombinant Tom40 inserted into liposomes was treated with trypsin (Fig. 1d, lower panel, lanes 2, 3). The remaining fragment of Tom40 was protected by the membranes, as, after lysis of mitochondria or liposomes with Triton X-100, it was digested by trypsin (Fig. 1d, lane 4). Antibodies directed against the amino-terminal or carboxy-terminal regions (12 terminal amino-acid residues in each region) of Tom40 showed that the C-terminal epitope of mitochondrial Tom40 was retained in the fragment (Fig. 1d, upper panel, lanes 6, 7), whereas the N-terminal epitope was removed by trypsin (Fig. 1d, upper panel, lane 10). The same topology of Tom40 was seen in the proteo-liposomes (Fig. 1d, lower panel, lanes 6, 7, 10). A quantitative assessment of the fragmentation of Tom40 in mitochondria and liposomes (including a consideration of the higher resistance of mitochondrial membranes to trypsin) indicates that at least 50–60% of the recombinant Tom40 was inserted into liposomes with the correct orientation (another fraction of recombinant Tom40 was fully digested by trypsin even in the presence of intact liposome membranes, probably representing Tom40 associated with the liposome preparation but not inserted into the membrane). Indeed, a preprotein specifically affected the channel activity only when added to the *cis* side of the membrane (that is, to the N terminus of Tom40) (see below). We conclude that the recombinant Tom40 molecules that were functionally incorporated into liposomes were asymmetrically inserted with the same relative orientation as in the outer mitochondrial membrane.

To determine whether reconstituted Tom40 could form a channel, we performed electrophysiological studies. After fusion of Tom40-containing liposomes with a planar bilayer, single-channel currents could indeed be measured (Fig. 2a). At membrane potentials below 100 mV, the channels were mainly open and we observed brief, flickering closures of the channel. When examined at higher time resolution (10 kHz), the current recordings showed subconductance levels (lower trace, Fig. 2a). The current traces show that

direct transitions between the two main conductance levels occur, indicating that these levels are (sub)conductant states of the same channel. At more negative membrane potentials, more subconductance levels were found (details not shown). With 250 mM KCl buffer on both sides of the membrane, the fully open channel showed a linear current-voltage relationship with slope conductance of $\Lambda = 360 \pm 3$ pS, whereas a value of $\Lambda = 150 \pm 2$ pS was observed for the most frequent subconductance level (Fig. 2b). In asymmetric buffers (250/20 mM KCl) the reversal potential was $E_{rev} = +40 \pm 1.3$ mV for the open single channel ($n > 60$; details not shown). Accordingly, the channel showed a selectivity for cations over anions ($P_{K^+} : P_{Cl^-} \approx 8 : 1$). The relative permeabilities for different cations calculated from the corresponding reversal potentials (CS^+ , 1.1; K^+ , 1.0; Na^+ , 0.7; Li^+ , 0.5; tetraethylammonium (TEA^+), 0.25; tetrabutylammonium (TBA^+), 0.12) roughly reflect the relative mobility of these ions in water. Therefore these ions are likely to cross through the sufficiently wide channel without interaction with it.

Figure 2c shows the voltage dependence of the channel open probability. At a membrane potential of 0 mV, the channel was completely open, whereas it closed symmetrically at positive or negative membrane potentials. When the concentrations of permeant ions on both sides were raised, the channel conductance increased asymptotically to a saturation value of $\Lambda \approx 3.58 \pm 0.35$ nS (Fig. 2d). Using the model that the Tom40 channel is a uniform cylinder filled with a solution of bulk medium resistivity, its diameter can be calculated according to refs 12, 13; when using $\Lambda = 360$ pS (symmetrical 250 mM KCl; assumed length of 50 Å), a diameter of 12 Å is obtained. However, on the basis of calculations in ref. 14, which assume that the resistivity in the channel is five

times the bulk resistivity, a pore diameter of 26 Å is obtained. An experimental assessment of the pore diameter using cations of various size showed that the Tom40 channels were permeable to TEA^+ , TBA^+ and different polyamines (cadaverine, spermine), whereas the polycation poly(L-Lysin)₁₀₀₀ partially blocked the channel. Using the polymer-exclusion method¹⁵, we obtained a pore diameter of 22 ± 1.4 Å.

When we incubated Tom40 with antibodies against the N-terminal epitope of Tom40, the proteoliposomes were precipitated and no further bilayer fusions were seen. The calculated number of Tom40 molecules in the liposomes correlated well with the number of active channels observed after fusion of the proteoliposomes to the planar bilayer. Together with the high purity of the Tom40 preparation, these data exclude the possibility that a contaminating protein was responsible for the channel activity. Moreover, a polyhistidine-tagged version of Tom40, purified by an independent procedure, yielded channel activities that were indistinguishable from those obtained with authentic Tom40. The basic properties of the reconstituted Tom40 channel (selectivity, conductance and open probability) were highly reproducible in more than 500 different experiments. We conclude that Tom40 forms a cation-selective high-conductance channel with multiple conductance states.

We isolated outer membrane vesicles^{16,17} and fused them to liposomes and then to a planar bilayer. We identified a channel (Fig. 2e, f) with the same conductance ($\Lambda = 224 \pm 10$ pS slope conductance, 250/20 mM KCl), cation selectivity ($E_{rev} = +40$ mV, 250/20 mM KCl, $P_{K^+} : P_{Cl^-} = 8 : 1$), and voltage dependence as the channel formed by purified Tom40. The open probability of the outer-membrane-integrated channel was about three- to fourfold

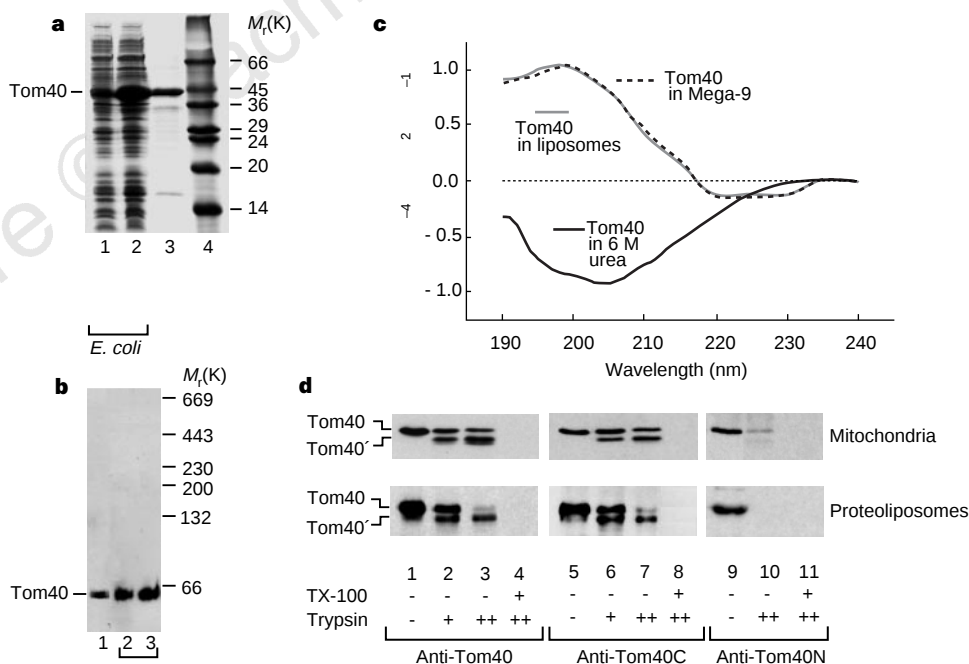


Figure 1 Expression of yeast Tom40 and its reconstitution into liposomes. **a**, Tom40 was analysed by Coomassie brilliant blue R250-stained SDS-PAGE. Lane 1, lysate from non-induced *E. coli* cells; lane 2, lysate from induced *E. coli* cells; lane 3, purified Tom40; lane 4, molecular size markers. **b**, Analysis of Tom40 in Mega-9 (1 µg in lane 2; 2 µg in lane 3) and in Mega-9-lysed mitochondria (200 µg in lane 1) by blue native gel electrophoresis and subsequent immunodecoration with antibodies against Tom40. **c**, Circular dichroism spectra of Tom40 in 6 M urea or 60 mM Mega-9 or reconstituted into liposomes. **d**, Orientation of Tom40 in mitochondria and liposomes. Treatment with trypsin was done as described in

Methods (as controls, mitochondria or liposomes were lysed with Triton X-100 (TX-100) before the protease treatment). Immunodecoration was done with antibodies generated against whole Tom40 (left panel), the C-terminal epitope (middle panel) and the N-terminal epitope (right panel). The trypsin-generated fragment of Tom40 is termed Tom40'. The treatment with trypsin reduces the affinity of anti-Tom40N to Tom40 (compare lane 10 with lanes 3, 7), probably because of removal of a small number of N-terminal amino acids (not changing the apparent gel mobility).

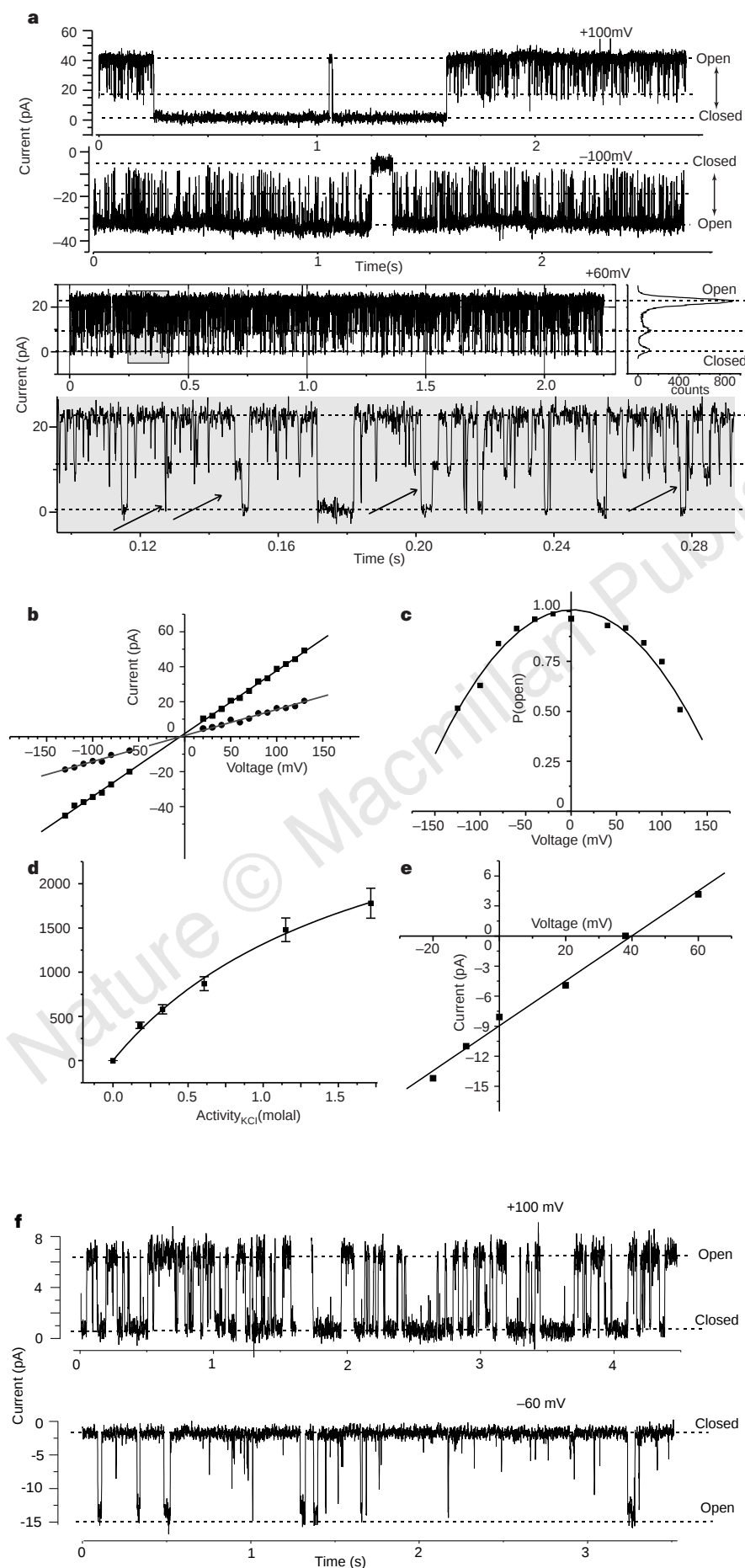
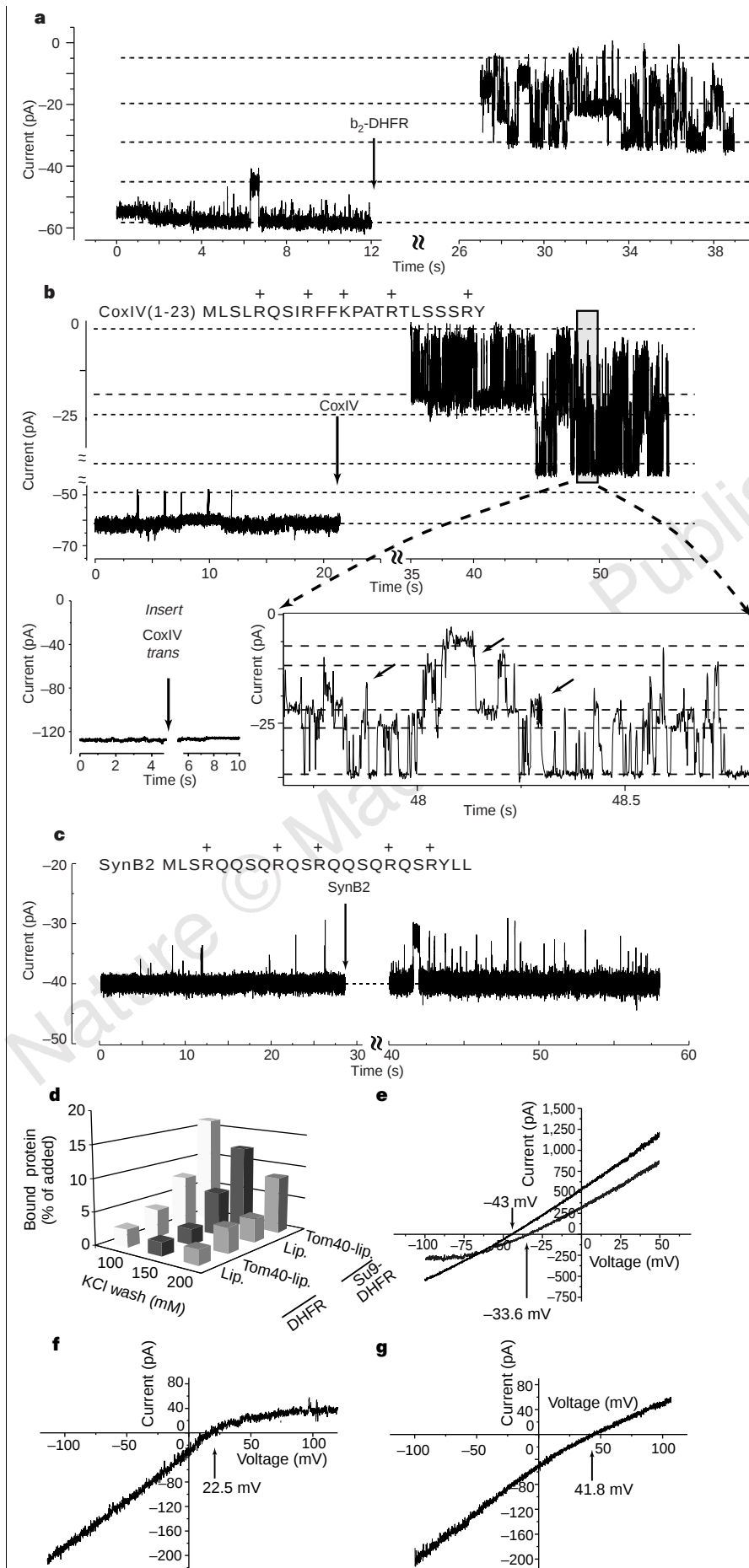


Figure 2 Tom40 forms a high-conductance cation-selective channel. **a**, Current traces from a bilayer containing one active Tom40 channel at different membrane potentials with 250 mM KCl, 10 mM CaCl₂ and 10 mM MOPS/Tris (pH 7.2) on both sides of the membrane. Applied membrane potentials are indicated. The lowest trace shows a timescale-expanded current recording from the trace above with high time resolution (10 kHz). Direct transitions between the different subconductance levels are marked with arrows. **b**, Current-voltage relationship of the fully open single channel (squares) and the most frequent subconductance level (circles) with 250 mM KCl on both sides of the membrane (averages of six independent bilayers, with standard errors of the means <4%). **c**, Voltage dependence of the probability for the Tom40 channel being in any of its open states. To approach equilibrium of channel gating with respect to the applied membrane potential voltages were applied for 5 min but only the current recordings of the last min were used to calculate P_{open} from the amplitude histograms (averages of at least three independent bilayers, with standard errors of the means <8% of the values). **d**, Saturation of the Tom40 conductance with symmetrically increasing KCl activities on both sides of the membrane. Bath solutions were changed by perfusion (three times the volume of the chamber). Data are averages of slope conductances from four different bilayers. **e, f**, Isolated outer mitochondrial membranes^{16,17} were fused with preformed liposomes by freeze-thawing and subsequently sonicated in a water bath. Such liposomes were fused with the bilayer and current recordings were performed in 250 mM KCl, 10 mM CaCl₂/20 mM KCl (*cis/trans* chamber) and 10 mM MOPS/Tris (pH 7.2). **e**, Current-voltage relationship of the outer mitochondrial membrane cation-selective channel (data points are averages of at least four independent bilayers, with standard errors of the means <3–5% of the values). **f**, Single-channel current recordings from a bilayer containing a single copy of the outer mitochondrial membrane cation-selective channel.

Figure 3 The Tom40 channel is sensitive to and transports a presequence peptide. **a–c**, Current traces from bilayers containing active Tom40 channels before and after addition of preprotein/peptides (b_2 -DHFR, CoxIV, SynB2) bathed in 250 mM KCl, 10 mM CaCl_2 /20 mM KCl (*cis/trans* chamber) and 10 mM MOPS/Tris (pH 7.2); holding potential $V_h = 0$ mV. After addition of the preprotein/peptides, the bath solution was vigorously stirred for 10 s. Dotted horizontal lines indicate the levels of the main conductances. **a**, Addition of 15 μM b_2 -DHFR to the *cis* compartment; average current reduction 52%. **b**, Addition of 15 μM CoxIV(1–23) to the *cis* compartment; average current reduction 51% ($n = 10$, s.e.m. 3%). Inset, addition of CoxIV(1–23) to the *trans* compartment (no current reduction). **c**, Addition of 15 μM SynB2 (no current reduction). The amino acid sequences of CoxIV(1–23) and SynB2 are shown in the single-letter code, with positive residues indicated. **d**, ^{35}S -labelled Su9-DHFR and DHFR were incubated with liposomes (lip.) containing reconstituted Tom40 or with an equal amount of liposomes lacking Tom40. After washing of the liposomes with KCl as indicated, the amount of bound Su9-DHFR and DHFR was quantified by digital autoradiography. The total amount of Su9-DHFR or DHFR added was set to 100%. After subtraction of protein bound to Tom40-free liposomes, 6.6% of Su9-DHFR and 1.1% of DHFR bound to Tom40-containing liposomes (on average). **e–g**, Translocation of CoxIV(1–23). Here voltages refer to the *cis* compartment. Voltage ramps ($\Delta V = 1 \text{ mV s}^{-1}$) were applied across bilayers containing multiple copies of the active Tom40 channel. **e**, Recording from a bilayer containing about 60 active copies of the Tom40 channel in asymmetrical 250 mM/20 mM KCl, 10 mM MOPS/Tris (pH 7.2) (*cis/trans*); upper trace, control; lower trace, addition of 50 μM CoxIV to the *trans* compartment. **f, g**, Recording from a bilayer containing about eight active copies of the Tom40 channel in asymmetrical 20 mM/250 mM KCl, 10 mM MOPS/Tris (pH 7.2) (*cis/trans*). **f**, Addition of 10 μM CoxIV(1–23) to the *cis* compartment. **g**, Control.



lower than that of purified Tom40, indicating that the presence of other Tom proteins may exert a regulatory effect on the gating of the channel.

We determined whether the channel formed by reconstituted Tom40 responded to the presence of a mitochondrial preprotein. We added a fusion protein containing part of the precursor of cytochrome b_2 together with dihydrofolate reductase (b_2 -DHFR)¹⁸ to the planar bilayer. This fusion protein strongly reduced the channel open probability and concomitantly increased the frequency of channel gating (Fig. 3a). This markedly reduced the total current passing through the Tom40 channel. This effect was only seen when the preprotein was added to the *cis* side of the planar bilayer (N terminus of Tom40). The effect was reversed when the preprotein was removed by perfusion of the chamber bath solution.

To determine the specificity of this inhibition by a preprotein, we compared the effects of two synthetic peptides on the channel activity, namely, a peptide corresponding to the mitochondrial-targeting signal of cytochrome oxidase subunit IV (CoxIV(1–23)) (Fig. 3b), and a related peptide (SynB2) with the same number of positively charged residues (Fig. 3c) that was, however, unable to direct the import of mitochondrial preproteins^{19,20}. CoxIV(1–23) efficiently reduced the channel open probability and increased the frequency of channel gating in the same way as did b_2 -DHFR (Fig. 3b), whereas SynB2 did not (Fig. 3c). The inhibitory effect of CoxIV(1–23) was, again, only seen when the peptide was added to the *cis* compartment (Fig. 3b). By adding b_2 -DHFR or CoxIV(1–23), we reduced currents in both directions, whereas currents were not reduced on addition of 15 μ M SynB2. The reduction in the mean currents was nonlinearly dose-dependent and 15 μ M b_2 -DHFR and CoxIV(1–23) led to roughly half-maximal inhibition. A large conductance channel, PSC, with some preference for cations has been localized to the outer membrane and shown to be blocked by the presequence peptide CoxIV and the non-mitochondrial basic peptide dynorphin B (ref. 6). The Tom40 channel, however, was not affected by dynorphin B (at 15 μ M; $n = 5$), supporting the conclusion that Tom40 interacts specifically with presequences, but not with other basic peptides.

Crosslinking studies in intact mitochondria showed that there is a specific interaction between Tom40 and mitochondrial presequences²¹. To test biochemically whether reconstituted Tom40 interacts with a mitochondrial preprotein, we incubated a fusion protein, consisting of the presequence of F_0 -ATPase subunit 9 (Su9) and the non-mitochondrial passenger protein dihydrofolate reductase (DHFR), with Tom40-containing liposomes or Tom40-free liposomes. A large excess of bovine serum albumin was included in the binding assay to exclude the possibility that unspecific protein–protein interactions were measured. After washing the liposomes with KCl to reduce the association of Su9–DHFR with the phospholipid headgroups, we observed a significant association of Su9–DHFR with the Tom40-containing liposomes (Fig. 3d; 200 mM KCl). The binding was mediated by the presequence, as DHFR alone showed only background levels of binding (Fig. 3d).

The complete translocation of presequence-containing preproteins across the outer mitochondrial membrane requires the import-driving system of the inner membrane (Tim–Hsp70 (heat-shock protein 70) machinery). When using isolated outer membrane vesicles, only translocation of the presequence could be monitored²². We therefore determined whether the purified Tom40 could also translocate presequences. We measured the current–voltage relationship of Tom40-containing bilayers in the presence of 5–100 μ M CoxIV(1–23) or 5–100 μ M SynB2 on either side of the membrane. When 10 μ M CoxIV(1–23) was added to the *cis* compartment containing 20 mM KCl (250 mM KCl *trans*), a pronounced change in the reversal potential was observed ($\Delta E = 18 \pm 2$ mV; $n = 11$) (Fig. 3f, g), and at positive potentials, which force the fivefold positively charged polypeptide towards the

trans compartment, a drastic decrease in the current was observed (Fig. 3f, g). When the experiments were done in the opposite direction, a significantly smaller change in the reversal potential occurred (Fig. 3e) ($\Delta E = 9 \pm 2.5$ mV; $n = 9$) and, concomitantly, the reduction in current was less pronounced (Fig. 3e).

These results show that CoxIV(1–23) can interact with the channel pore from both sides; however, its affinity for the pore is about one order of magnitude higher when it is on the *cis* side of the pore indicating that the presequence may interact with negatively charged groups of the pore. As $<10 \mu$ M of CoxIV(1–23) on the *cis* side (low salt) changed the reversal potential markedly, it is unlikely that this effect is due only to the permeation of the peptide. As CoxIV(1–23) did not completely block the channel pore, we estimate, from comparison with the current–voltage relationship of spermine, that the presequence peptide can permeate the channel at significant rates (up to 10^3 – 10^4 molecules per s at a driving force of 100 mV). After addition of SynB2 (50 μ M) to either compartment ($n = 5$ each side), $\Delta E = 10 \pm 3$ mV. Thus, only CoxIV(1–23) shows high affinity for the pore (*cis* side). Moreover, the suppression of the current by SynB2 was about 10–50-fold more pronounced than that by CoxIV(1–23), indicating that SynB2 is transported at a considerably slower rate than the presequence peptide. The studies using low membrane potential (Fig. 3b) show that the presequence peptide selectively binds to Tom40 from the *cis* side and thereby alters the gating of the channel, whereas it is rapidly translocated through the channel when driven by higher voltage (Fig. 3f).

Tom40 is, to our knowledge, the first membrane protein of the mitochondrial protein import machinery that has been purified and reconstituted in a functional form. It forms a hydrophilic cation-selective channel. The size of the channel diameter (~ 22 Å) indicates that typical folded proteins cannot permeate; precursor polypeptides may pass through in an α -helical conformation or in an extended conformation. A piece of double-stranded DNA (diameter ~ 20 Å) covalently linked to a preprotein has been imported into mitochondria²³. The size of the Tom40 channel seems to be sufficient to allow passage of the DNA. Recently, a Tom complex containing about seven different subunits was isolated from *Neurospora crassa* mitochondria. Negative stain electron microscopy showed stain-filled pits with an apparent diameter of ~ 20 Å that may represent import pores²⁴. We propose that the pores were formed by Tom40. In the protein-import machinery of the chloroplast outer-envelope membrane, Toc75 (OEP75) was identified as the pore-forming subunit¹³. Toc75 has mainly β -sheet structure, like Tom40, but Toc75 and Tom40 share no sequence homology and activation of the Toc75 channel depends completely on a membrane potential. The calculated channel diameter for Toc75 is smaller (~ 10 Å) (ref. 13) than that of Tom40, whereas endoplasmic reticulum protein-transport channel, formed by the hetero-oligomeric Sec61p complex, has a calculated diameter of 20–60 Å (ref. 25).

Tom40 was identified as an outer-membrane protein that is near to a preprotein in transit²⁶. The mature part²⁶ and the presequence²¹ of a preprotein were crosslinked to Tom40, and Tom40 was shown to be an essential component of the multisubunit Tom machinery^{8,17,27,28}. In addition, four other Tom proteins, Tom70, Tom22, Tom20 and Tom5, have been found to be crosslinked to preproteins^{16,17,28,29}. Despite the complexity of the preprotein translocation of the mitochondrial outer membrane, it is now possible to assign individual functions to these Tom proteins: Tom70, Tom20 and Tom22 have receptor/binding functions, as shown using purified cytosolic and intermembrane space domains^{20,30}; Tom5 is a mediator between receptors and the import pore¹⁷; and Tom40 forms the hydrophilic channel of the general import pore for preproteins. The affinity of Tom40 for mitochondrial-targeting sequences added to the *cis* side is close to the range of affinities seen with several Tom receptors²⁰. We suggest that Tom40 is part of the proposed chain of multiple interaction sites for preprotein

transport across the outer membrane^{17,29,30}. The asymmetric affinity of Tom40 for targeting sequences favours transmembrane transport in the *cis* to *trans* direction and renders Tom40 an important component that contributes to the directionality of mitochondrial-preprotein import. □

Methods

Expression and reconstitution of Tom40 in liposomes. A 1,164-base-pair genomic fragment, encompassing the open reading frame of *TOM40*, was amplified with *Pfu*-polymerase, subcloned into the vector pET11a and transformed into *E. coli* BL21(DE3). After induction with isopropyl- β -D(-)-thiogalactopyranoside, Tom40 was purified by a method adapted from ref. 10. Inclusion bodies were isolated, extensively washed in buffer containing Triton X-100 and solubilized in urea buffer (8 M urea, 50 mM Tris/HCl pH 8.0, 1 mM EDTA and 100 mM dithiothreitol (DTT)). The pH was gradually lowered to 6.5, followed by centrifugation at 50,000g, and the supernatant containing purified Tom40 was dialysed against 6 M urea, 50 mM Tris/HCl pH 8.0, 1 mM EDTA and 2 mM DTT. Tom40 in urea was diluted tenfold in Mega-9 buffer (80 mM Mega-9, 10 mM MOPS/Tris pH 7.0, 1 mM EDTA and 100 mM DTT) and analysed by blue native gel electrophoresis¹⁸.

Purified Tom40 was resuspended in 80 mM Mega-9, 6 M urea and 10 mM MOPS/Tris pH 7.0 and reconstituted in liposomes by the dialysis technique. Briefly, small unilamellar liposomes were obtained from purified azolectin (Sigma, type IV S) as described¹³ and were lysed in Mega-9. Tom40 was diluted ~50-fold into the mixture and proteoliposomes were then formed by dialysis. The proteoliposomes contained ~10–20 molecules of Tom40 per liposome. Circular dichroism spectra were recorded using a Jasco-J-600A spectropolarimeter as described¹³. Samples were adjusted to the same protein concentration ($100 \pm 30 \mu\text{g}$) and spectra were averaged ($n = 30$) to improve the signal:noise ratio. For digestion with trypsin, proteoliposomes were concentrated by discontinuous sucrose or Ficoll density gradient centrifugation and treated with $1\text{--}5 \mu\text{g ml}^{-1}$ trypsin. Mitochondria were treated with $10\text{--}50 \mu\text{g ml}^{-1}$ trypsin (the presence of other Tom proteins requires a higher concentration of trypsin in order to obtain access to Tom40 (refs 16, 17)). The preprotein Su9-DHFR and DHFR were synthesized in rabbit reticulocyte lysate in the presence of [³⁵S]methionine/cysteine¹⁶ and incubated with Tom40-containing liposomes and Tom40-free liposomes in buffer containing 20 mM KCl and 2% (w/v) BSA, followed by addition of further KCl (concentration as in wash buffer). The liposomes were isolated by centrifugation and washed in KCl (100–200 mM)-containing buffer.

Electrophysiological measurements. Planar lipid bilayers were produced using the painting technique¹³. The resulting bilayers had a typical capacitance of $\sim 0.5 \mu\text{F cm}^{-2}$ and a resistance of $>100 \text{ G}\Omega$ (root mean square $\approx 1 \text{ pA}$ at 5 kHz bandwidth). After a stable bilayer was formed in symmetrical solutions of 20 mM KCl, 10 mM MOPS/Tris pH 7.0, the solution of the *cis* chamber was changed to asymmetrical concentrations (*cis* chamber: 250 mM KCl, 10 mM CaCl_2 , 10 mM MOPS/Tris pH 7.0) by adding concentrated solutions of KCl and CaCl_2 . The liposomes were added to the *cis* compartment directly below the bilayer through the tip of a micropipette to allow the flow of the liposomes across the bilayer. If necessary, the solution in the *cis* chamber was stirred to promote fusion. After fusion, the electrolytes in both compartments were changed to the final composition by perfusion. The Ag/AgCl electrodes were connected to the chambers through 2 M KCl–agar bridges. The electrode of the *trans* compartment was directly connected to the headstage of a current amplifier (EPC 7, List Medical). Reported membrane potentials are referred to the *trans* compartment. The amplified currents were recorded on a modified DAT recorder digitized at sampling intervals of 0.05–0.2 ms.

Received 8 June; accepted 7 July 1998.

- Schatz, G. & Dobberstein, B. Common principles of protein translocation across membranes. *Science* **271**, 1519–1526 (1996).
- Neupert, W. Protein import into mitochondria. *Annu. Rev. Biochem.* **66**, 863–917 (1997).
- Pfanner, N., Craig, E. A. & Hönlinger, A. Mitochondrial preprotein translocase. *Annu. Rev. Cell Dev. Biol.* **13**, 25–51 (1997).
- Moran, O., Sandri, G., Panfil, E., Stühmer, W. & Sorgato, C. Electrophysiological characterization of contact sites in brain mitochondria. *J. Biol. Chem.* **265**, 908–913 (1990).
- Lohret, T. A., Jensen, R. E. & Kinnally, K. W. Tim23, a protein import component of the mitochondrial inner membrane, is required for normal activity of the multiple conductance channel, MCC. *J. Cell Biol.* **137**, 377–386 (1997).
- Juin, P., Thieffry, M., Henry, J.-P. & Vallette, F. M. Relationship between the peptide-sensitive channel

- and the mitochondrial outer membrane protein translocation machinery. *J. Biol. Chem.* **272**, 6044–6050 (1997).
- Dihanich, M., Schmidt, A., Oppliger, W. & Benz, R. Identification of a new pore in the mitochondrial outer membrane of a porin-deficient yeast mutant. *Eur. J. Biochem.* **181**, 703–708 (1989).
 - Baker, K. P., Schaniel, A., Vestweber, D. & Schatz, G. A yeast mitochondrial outer membrane protein essential for protein import and cell viability. *Nature* **348**, 605–609 (1990).
 - Mannella, C. A., Neuwald, A. F. & Lawrence, C. E. Detection of likely transmembrane β -strand regions in sequences of mitochondrial pore proteins using the Gibbs sampler. *J. Bioenerg. Biomembr.* **28**, 163–169 (1996).
 - Schmid, B., Krömer, M. & Schulz, G. E. Expression of porin from *Rhodospseudomonas blastica* in *Escherichia coli* inclusion bodies and folding into exact native structure. *FEBS Lett.* **381**, 111–114 (1996).
 - Sreerama, N. & Woody, R. W. Protein secondary structure from circular dichroism spectroscopy: combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods. *J. Mol. Biol.* **242**, 497–507 (1994).
 - Hille, B. in *Ionic Channels of Excitable Membranes*. Ch. 9, 11 (Sinauer, Sunderland, Massachusetts, 1992).
 - Hinnah, S., Hill, K., Wagner, R., Schlicher, T. & Soll, J. Reconstitution of a chloroplast protein import channel. *EMBO J.* **16**, 7351–7360 (1997).
 - Smart, O. S., Breed, J., Smith, G. R. & Sansom, M. S. A novel method for structure-based prediction of ion channel conductance properties. *Biophys. J.* **72**, 1109–1126 (1997).
 - Krasilnikov, O. V., Sabirov, R. Z., Ternovsky, V. L., Merzliak, P. G. & Muratkhodjaev, J. N. A simple method for the determination of the pore radius of ion channels in planar lipid bilayer membranes. *FEMS Microbiol. Immunol.* **5**, 93–100 (1992).
 - Alconada, A., Gärtner, F., Hönlinger, A., Kübrich, M. & Pfanner, N. Mitochondrial receptor complex from *Neurospora crassa* and *Saccharomyces cerevisiae*. *Methods Enzymol.* **260**, 263–286 (1995).
 - Dietmeier, K. et al. Tom5 functionally links mitochondrial preprotein receptors to the general import pore. *Nature* **388**, 195–200 (1997).
 - Dekker, P. J. T. et al. The Tim core complex defines the number of mitochondrial translocation contact sites and can hold arrested preproteins in the absence of matrix Hsp70-Tim44. *EMBO J.* **16**, 5408–5419 (1997).
 - Allison, D. S. & Schatz, G. Artificial mitochondrial presequences. *Proc. Natl Acad. Sci. USA* **83**, 9011–9015 (1986).
 - Brix, J., Dietmeier, K. & Pfanner, N. Differential recognition of preproteins by the purified cytosolic domains of the mitochondrial import receptors Tom20, Tom22, and Tom70. *J. Biol. Chem.* **272**, 20730–20735 (1997).
 - Rapaport, D., Neupert, W. & Lill, R. Mitochondrial protein import: Tom40 plays a major role in targeting and translocation of preproteins by forming a specific binding site for the presequence. *J. Biol. Chem.* **272**, 18725–18731 (1997).
 - Mayer, A., Neupert, W. & Lill, R. Mitochondrial protein import: reversible binding of the presequence at the *trans* side of the outer membrane drives partial translocation and unfolding. *Cell* **80**, 127–137 (1995).
 - Vestweber, D. & Schatz, G. DNA-protein conjugates can enter mitochondria via the protein import pathway. *Nature* **338**, 170–172 (1989).
 - Künkele, K.-P. et al. The preprotein translocation channel of the outer membrane of mitochondria. *Cell* **93**, 1009–1019 (1998).
 - Matlack, K. E. S., Mothes, W. & Rapoport, T. A. Protein translocation: tunnel vision. *Cell* **92**, 381–390 (1998).
 - Vestweber, D., Brunner, J., Baker, A. & Schatz, G. A 24K outer-membrane protein is a component of the yeast mitochondrial protein import site. *Nature* **341**, 205–209 (1989).
 - Kiebler, M. et al. Identification of a mitochondrial receptor complex required for recognition and membrane insertion of precursor proteins. *Nature* **348**, 610–616 (1990).
 - Söllner, T. et al. Mapping of the protein import machinery in the mitochondrial outer membrane by crosslinking of translocation intermediates. *Nature* **355**, 84–87 (1992).
 - Hönlinger, A. et al. The mitochondrial receptor complex: Mom22 is essential for cell viability and directly interacts with preproteins. *Mol. Cell Biol.* **15**, 3382–3389 (1995).
 - Bolliger, L., Junne, T., Schatz, G. & Lithgow, T. Acidic receptor domains on both sides of the outer membrane mediate translocation of precursor proteins into yeast mitochondria. *EMBO J.* **14**, 6318–6326 (1995).

Acknowledgements. We thank P. J. T. Dekker, B. Guiard, M. Kiebler, W. Kühlbrandt, D. Mills, J. Brix, S. C. Hinnah and H. Martin for materials, advice and discussion. This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 388, the Fonds der Chemischen Industrie (N.P.), the Sonderforschungsbereich 171 (R.W.) and a longterm fellowship of the Alexander-von-Humboldt Foundation (M.T.R.).

Correspondence and requests for materials should be addressed to N.P. (e-mail: pfanner@ruf.uni-freiburg.de).

Drosophila CBP represses the transcription factor TCF to antagonize Wingless signalling

Lucas Waltzer & Mariann Bienz

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

T-cell factor (TCF), a high-mobility-group domain protein, is the transcription factor activated by Wnt/Wingless signalling^{1–4}. When signalling occurs, TCF binds to its coactivator, beta-catenin/Armadillo, and stimulates the transcription of the target genes of Wnt/Wingless by binding to TCF-responsive enhancers^{1,5}. Inappropriate activation of TCF in the colon epithe-

lium and other cells leads to cancer⁶⁻⁸. It is therefore desirable for unstimulated cells to have a negative control mechanism to keep TCF inactive. Here we report that *Drosophila* CREB-binding protein (dCBP)^{9,10} binds to dTCF. dCBP mutants show mild Wingless overactivation phenotypes in various tissues. Consistent with this, dCBP loss-of-function suppresses the effects of *armadillo* mutation. Moreover, our data show that dCBP acetylates a conserved lysine in the Armadillo-binding domain of dTCF, and that this acetylation lowers the affinity of Armadillo binding to dTCF. Although CBP is a coactivator of other transcription factors^{11,12}, our data show that CBP represses TCF.

CBP (also known as p300) is a transcriptional coactivator that is recruited to DNA by several transcription factors including CREB (for cAMP response-element-binding protein)¹¹ and c-Fos¹³. We used the yeast two-hybrid system to assess binding of dCBP to candidate DNA-binding proteins that recognize the minimal signal-responsive element in the *Ultrabithorax* (*Ubx*) midgut enhancer (for example, dTCF, dCREB and dFOS)^{14,15}. Unexpectedly, we found

that dCBP binds tightly to dTCF. dTCF binds to the dCBP2 domain (Fig. 1a), which is homologous to the mouse CBP2 (or C/H3) domain¹³. Binding between dCBP2 and dTCF was confirmed *in vitro* by a pull-down assay using a bacterially expressed dCBP2 fusion protein (glutathione S-transferase (GST)-dCBP2) and *in vitro* translated dTCF (Fig. 1b). This interaction is conserved between species, as dCBP2 also binds to mouse LEF-1 (Fig. 1b), a close relative of TCF, and mouse CBP2 binds to dTCF and LEF-1 (data not shown). Using *in vitro* translated dCBP2 and various domains of dTCF expressed as GST fusion proteins in bacteria, we found that the high-mobility group (HMG) domain of dTCF is sufficient to bind dCBP2 (Fig. 1c).

We then investigated whether *nejire* (*nej*) mutants, which lack dCBP function^{9,10}, show phenotypes related to those of Wingless-pathway mutants and, if so, whether they mimic hyperactivation of the pathway (as do *shaggy/zeste-white 3* (*sgg*) mutants) or lack of pathway activity (as do *wingless* or *armadillo* (*arm*) mutants)¹⁶. Examination of the strongest available *nej* allele, *nej*³, revealed that nearly all mutant embryos look comparatively normal and develop to the end of embryogenesis, but their midguts show various abnormalities. We stained these embryos with an antibody against Labial, a HOX protein encoded by an endodermal Wingless target gene¹⁷. A widening of Labial staining was common in the *nej* mutants (Fig. 2b), as were gaps in the staining (Fig. 3b, arrowheads), indicating the presence of non-staining cells between abnormally strongly staining cells. A similar phenotype is observed after Wingless or Armadillo overexpression, which stimulates *labial* at low signalling levels but represses *labial* at high levels (X. Yu and M.B., unpublished data)¹⁸. In contrast, *arm* mutants show uniformly reduced and shallow Labial staining¹⁹ (Fig. 3d). Clearly, the Labial staining pattern in *nej* mutants looks like that of *sgg* rather than *arm*. Consistent with this, the midgut constriction is present in *nej*³ mutants (Figs 2b, f and 3g, arrows) but not in *arm* mutants¹⁹ (Fig. 3i). Thus, in the endoderm, dCBP loss-of-function mimics overactivation of the Wingless pathway, implying that dCBP is a negative regulator of this pathway in this midgut cell layer.

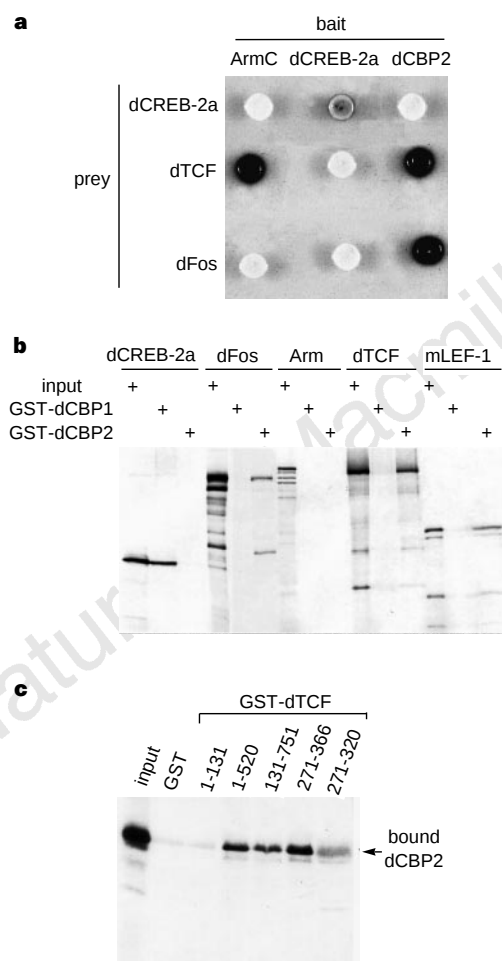


Figure 1 dCBP interacts with dTCF. **a**, Yeast two-hybrid assays. Culture spots from yeast transformed with different combinations of prey and bait plasmids. *lacZ* activity was tested on galactose plates (prey protein expressed; no *lacZ* activity was detected on control glucose plates on which prey expression is repressed). **b**, *In vitro* binding between dTCF and dCBP. Equimolar amounts of immobilized GST-dCBP1 or GST-dCBP2 were incubated with *in vitro* translated ³⁵S-labelled proteins as indicated (dCREB-2a was phosphorylated by protein kinase A before the binding assay¹¹); in the input lane, 20% of the total reaction was loaded. **c**, dCBP interacts with the HMG domain of dTCF. Bacterially expressed fragments of dTCF (remaining residues indicated) were purified and equimolar amounts of each fragment were incubated with *in vitro* translated ³⁵S-labelled dCBP2. Residues 271–366 span the HMG domain and constitute a minimal DNA-binding fragment (not shown). In the input lane, 20% of the total reaction was loaded.

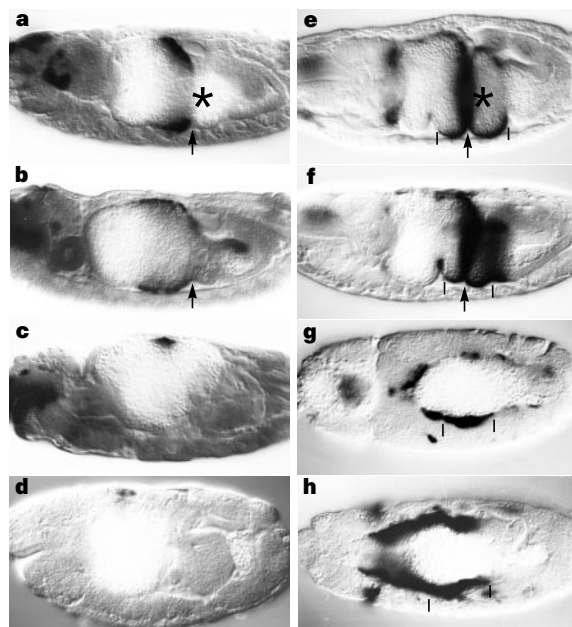


Figure 2 *nej* mutants mimic overactivation of the Wingless pathway. **a–h**, Side views of embryos 14–16 h old stained with Labial (**a–d**) or *lacZ* antibody (**e–h**) to visualize expression from the *Ubx* B midgut enhancer; **e, e**, wild type; **b, f**, *nej*³ hemizygotes; **c, g**, *nej*¹ hemizygotes from *nej*¹ germ lines; **d, h**, *sgg* hemizygotes from *sgg* germ lines. The midgut constriction is marked by an arrow, and the normal width of *Ubx* B expression by vertical bars; in the wild type, the position of the Wingless source is marked by an asterisk; anterior to the left, dorsal up.

We then assayed the effects of *nej* mutation on the activity of a dTCF target enhancer in the midgut¹. This enhancer from the *HOX* gene *Ubx* mediates β -galactosidase (*lacZ*) expression in the mesodermal cell layer of the midgut (Fig. 2e). In *arm* mutants, *lacZ* expression from this enhancer is drastically reduced¹ (Fig. 3i). In contrast, *lacZ* staining is significantly stronger in *nej*³ mutants, especially in cells distal to the Wingless source (Figs 2f and 3g). Moreover, *Ubx* enhancer activity is often de-repressed posteriorly beyond the normal expression domain (Figs 2f and 3g). If the TCF-binding site within the *Ubx* enhancer is mutated¹, we find no de-repression in *nej*³ mutants (data not shown). This implies that dCBP represses the *Ubx* midgut enhancer, and that this repression is mediated by the dTCF-binding site.

To see whether the mild gut phenotypes of *nej*³ mutants would be stronger in the absence of perduring maternal dCBP, we examined *nej*¹ mutant embryos derived from *nej*¹ mutant germ lines (*nej*³ mutant germ lines do not develop¹⁰). These mutant embryos show a range of phenotypes from relatively normal (similar to *nej*³ mutants; Fig. 2b, f) to severely abnormal (Fig. 2c, g). The most extreme cases show almost complete suppression of Labial staining (Fig. 2c), and striking de-repression of *Ubx* enhancer activity to the anterior end of the midgut (Fig. 2g). Again, these staining patterns resemble those seen in *sgg* mutants¹⁹ (Fig. 2d, h).

We looked for *sgg*-like phenotypes of *nej* in other developmental contexts, and found that embryonic cuticles of *nej*³ mutants frequently lack anterior denticles in individual denticle belts (Fig. 3l). This is similar to a mild Wingless overactivation phenotype^{20,21}.

We also examined dCBP function in the wing disc to find out

whether the mild *wingless* loss phenotype caused by overexpressed Cadherin-intra (Cad-I), which sequesters Armadillo²², depends on dCBP gene dosage. Overexpressed Cad-I causes large gaps in the wing margin (Fig. 3p) that are worse in *arm* heterozygotes²² (Fig. 3r). These margin defects are largely alleviated in *nej* heterozygotes (Fig. 3q), as in *sgg* heterozygotes (not shown; S. Greaves and J.-P. Vincent, personal communication). Once more, these results are consistent with dCBP's being a negative regulator of the Wingless pathway.

Given that *nej* mutants are *sgg*-like, we might expect non-additive effects in *sgg nej* double mutants. Embryos zygotically mutant for *sgg* show completely normal guts (Fig. 3a, f). However, *sgg nej* double mutants look more severe than *nej* mutants, with more Labial staining gaps than the *nej*³ mutants (Fig. 3c, arrowheads), and with stronger *lacZ* staining from the *Ubx* enhancer (Fig. 3h). This indicates a synergy between *nej* and *sgg* in the midgut.

We also generated *arm nej* double mutants to see whether *nej* would suppress *arm*. Zygotically mutant *arm* embryos show universally reduced Labial staining in the midgut (Fig. 3d), lack of the midgut constriction (Fig. 3l, open triangle) and substantial reduc-

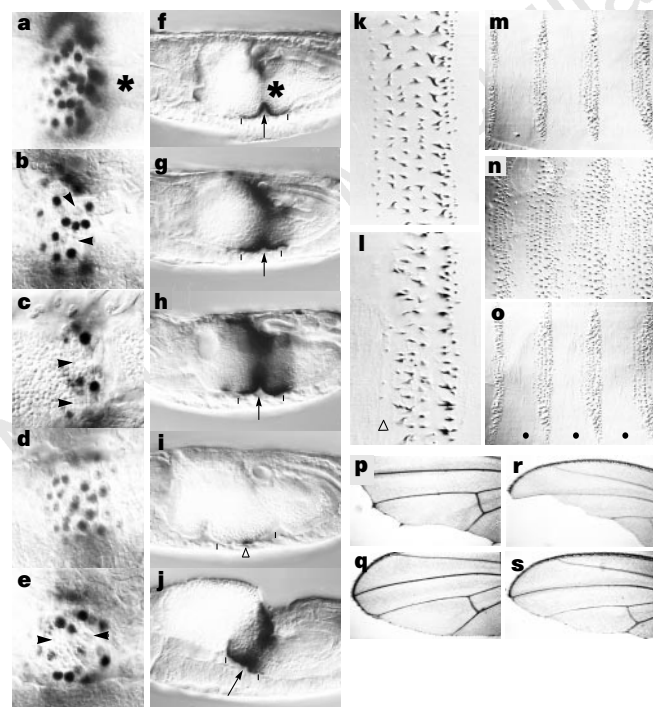


Figure 3 Genetic interactions between *nej* and Wingless pathway mutants. **a–j**, Side views of embryos 14–16 h old stained with Labial (**a–e**) or *lacZ* antibody (**f–j**) to visualize *Ubx* B expression; **a, f**, *sgg* hemizygote (indistinguishable from wild type); **b, g**, *nej*³ hemizygote; **c, h**, *sgg nej*³ hemizygote; **d, i**, *arm*^{XM19} hemizygote; **e, j**, *arm*^{XM19} *nej*³ hemizygote. Some gaps in Labial staining are indicated by arrowheads; lack of the midgut constriction in **i** is marked by a triangle; other symbols are as in Fig. 2. **k–o**, Ventral denticle belts from the larval abdomen; **k**, wild type; **l, m**, *nej*³ hemizygote; **n**, *arm*^{H8.6} hemizygote; **o**, *arm*^{H8.6} *nej*³ hemizygote; missing denticles in rows 1 and 2 in **l** are indicated by an open triangle, and the restoration of naked cuticle between denticle belts in **o** is marked by dots; anterior to the left. **p–s**, Distal wing parts from flies after Cad-I overexpression in the wing disc; **+/+** (**p**); *arm*^{XM19} **+/+** (**r**); *nej*³ **+/+** (**q**); *arm*^{XM19} *nej*³ **+/+** (**s**).

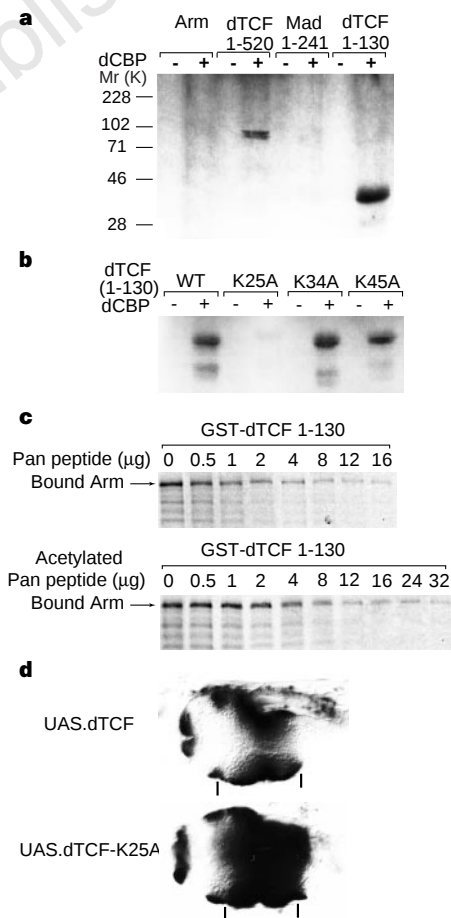


Figure 4 dCBP acetylates the Armadillo-binding domain of dTCF. **a**, Protein acetylation assay. Separated reaction products after incubation of purified protein with ¹⁴C-acetyl-CoA with (+) or without (–) dCBP HAT domain (Mad signifies a control protein). **b**, N-terminal dTCF domains with lysine substitutions, assayed for acetylation by dCBP and visualized as in **a, c**. Competition assays; wild-type or acetylated Pan peptide was added at increasing concentrations to a binding reaction between *in vitro* translated Armadillo and 5 μ g of immobilized GST-dTCF1–130; an average difference in competition of two- to threefold between the two peptides was consistently measured in three independent experiments. **d**, Enhanced transcriptional stimulation of the *Ubx* B enhancer by non-acetylatable dTCF-K25A compared to wild-type dTCF after overexpression in the mesoderm; embryos and symbols are as in Fig. 2.

tion of *Ubx* enhancer activity¹ (Fig. 3i). In contrast, *arm nej* double mutants resemble *nej* mutants by showing strong Labial staining, with some staining gaps (Fig. 3e), as well as increased and spatially de-repressed *Ubx* enhancer activity (Fig. 3j). Furthermore, the midgut constriction is restored in *arm nej* double mutants (Fig. 3j). We observed similar suppressions of *arm* or *wingless* loss in the larval cuticle and in the wing: naked cuticle stretches that are lacking in *arm* mutants (Fig. 3n) are restored in *arm nej* double mutants (Fig. 3o), and the wing margin that is notched in *arm* heterozygotes overexpressing Cad-I (Fig. 3p) is almost restored to a wild-type phenotype in *arm nej* heterozygotes (Fig. 3s). These double-mutant phenotypes indicate that *nej* acts downstream of, or in parallel to, *arm*, and imply that dCBP antagonizes Armadillo.

We investigated how dCBP might exert its repressor function at the molecular level. Binding of dTCF to DNA or Armadillo is not detectably affected by its binding to dCBP2. However, it has been shown that CBP and its functional homologue p300 acetylate histones^{23,24}, components of the basal transcription machinery²⁵, and a transcription factor²⁶. The histone acetyl transferase (HAT) domain of dCBP acetylates dTCF *in vitro* in its amino-terminal domain (Fig. 4a). Mutating individually each conserved lysine in this domain, we found that Lys 25 is the main, if not the only, residue in dTCF that is acetylated by dCBP (Fig. 4b).

Lys 25 is within the most conserved part of the dTCF N-terminus, and is six residues away from a point mutation in dTCF that reduces its binding to Armadillo³. Thus Armadillo binding to dTCF might interfere with the acetylation of Lys 25, and vice versa. Consistent with this, we found that pre-binding of Armadillo to the N terminus of dTCF abrogates acetylation (not shown). Conversely, the Pan peptide, a synthetic polypeptide spanning the putative Armadillo-binding site in dTCF, is consistently two- to threefold more effective in competing out binding of *in vitro* translated Armadillo to dTCF than a Pan peptide that is acetylated at Lys 25 (Fig. 4c). Thus acetylation of dTCF by dCBP lowers the binding affinity of dTCF to Armadillo *in vitro*.

To assess the functional importance of dTCF acetylation *in vivo*, we generated transformant embryos overexpressing a mutant form of dTCF whose Lys 25 was replaced by alanine (dTCF-K25A embryos). Overexpressed wild-type dTCF barely affects the activity of the *Ubx* midgut enhancer (Fig. 4d, top), but dTCF-K25A stimulates this enhancer significantly, sometimes beyond its normal expression domain (Fig. 4d, bottom). Evidently, a non-acetylatable dTCF is more active in stimulating transcription than the normal acetylatable dTCF.

Our biochemical and genetic analysis suggests that dCBP keeps dTCF inactive in the absence of Wingless signalling by acetylating its Armadillo-binding domain. At low levels of nuclear Armadillo, dCBP may block the recruitment of Armadillo by dTCF to Wingless-inducible genes. Upon signalling, Armadillo levels increase substantially²⁷, which would favour binding of Armadillo to (transiently) unacetylated dTCF (that is, we envisage a competing deacetylase). We therefore propose that the acetylation block imposed on dTCF by dCBP is overcome by a high concentration of Armadillo. We predict that the balance between Armadillo binding to dTCF and CBP acetylating it is particularly critical in cells that are near a low signalling threshold spatially or temporally.

Our results show that CBP can repress a transcriptional activator, TCF. This is in contrast to CBP's well-established role as a transcriptional coactivator^{12,23–25}. CBP evidently has several functions, and the previous analysis in *Drosophila* supports activating functions of dCBP^{9,10}, in addition to its repressive function shown here.

Finally, our model predicts that malfunction of CBP in unstimulated cells could result in the inappropriate activation of Wnt target genes by TCF and low levels of β -catenin. Indeed, point mutations have been found in the HAT and CBP2 domains of the remaining p300 gene in colon and gastric carcinomas that have lost one copy of this gene, indicating that p300/CBP has a tumor-suppressor

function²⁸. This is consistent with our data showing that dCBP represses dTCF to antagonize Armadillo-mediated activation of Wingless target genes. □

Methods

Drosophila strains and phenotypic analysis. The following mutant alleles or transformants were used: *nej*³, *nej*¹ (refs 9, 10), *arm*^{XM19}, *arm*^{H8.6} (ref. 27); *zw3*^{M11} (ref. 20); *Ubx* B, B4 (ref. 1); engrailed.GAL4 UAS.Cad-I (ref. 22); B24.GAL4 (ref. 29); UAS.dTCF (ref. 2). We observed only a few percent of *nej* mutant embryos that showed the twisted phenotype after which the mutation was named⁹. Double-mutant chromosomes (*arm*^{XM19} *nej*³, *arm*^{H8.6} *nej*³, *zw3*^{M11} *nej*³) were generated by standard methods and tested for presence of mutant alleles with Y-linked duplications of *arm* or *zw3* (provided by M. Pfeifer and P. Simpson). All mutants were identified by using 'blue balancers'. Germline clones were generated as described^{10,19}. *arm*^{H8.6} mutants were left to develop at 29°C for the midgut analysis (not shown). Antibody staining of embryos with *lacZ* (Promega) or Labial antibody¹⁵, and analysis of cuticles and wings, was done as described^{12,21}. Transformants of UAS.dTCF-K25A were established by standard methods, and four independent lines were compared with two independent UAS.dTCF lines; expression levels of wild-type and mutant dTCFs in the different lines were similar, as judged by antibody staining. The midgut phenotypes of *nej* mutants described cannot be accounted for by lack-of-function of *cubitus interruptus* or *dorsal*^{9,10} because these conditions result in less Labial staining or less mesoderm, respectively (not shown). Finally, the reported suppression of the wing margin defects in *Ci*^D by *nej* heterozygosity⁹ can be explained by our Cad-I results (Fig. 3) and the fact that *Ci*^D is also a dTCF loss-of-function allele^{2,3}.

Yeast two-hybrid assays. Interaction assays in yeast were done as described¹. Full-length dTCF, dFos and dCREB-2a (see ref. 14) were inserted by polymerase chain reaction (PCR) cloning into the prey vector pJG4-5. Similarly, full-length dCREB-2a, ArmC (residues 156–690) or dCBP2 (residues 2240–2507) were inserted into the bait vector pLex202.

In vitro binding assays. Various domains of dCBP (dCBP1, residues 895–1050; dCBP2, see above) and dTCF were cloned into pGEX-2T (Pharmacia) using PCR or standard cloning procedures. To express full-length Armadillo, pGEX3X-Arm was used¹. GST fusion proteins were produced in *Escherichia coli* BL21 and purified by affinity chromatography on GST–agarose beads. *In vitro* translated ³⁵S-methionine-labelled proteins were incubated with immobilized GST fusion proteins for 1 h at 4°C in buffer A (containing (in mM): 20 HEPES, pH 7.9, 200 NaCl, 1 EDTA, 1 MgCl₂, 1 DTT, 0.5% NP40) with 1 mg ml⁻¹ BSA. After extensive washing in buffer A, bound protein was eluted by boiling in SDS-loading buffer, resolved by SDS–PAGE and visualized by autoradiography. Synthetic wild-type and K25-acetylated Pan peptides (residues 19–49) were used to compete for binding of GST–dTCF1–130 to *in vitro* translated Armadillo; peptide concentrations were determined (at ~10% accuracy) by standard methods.

Protein acetyltransferase assay. Acetylation of protein was performed as described²⁴. Typically, 100 ng of the dCBP (residues 1675–2510) or the mouse CBP HAT domain (residues 1099–1758; provided by T. Kouzarides) and 2.5–5 μ g of substrate protein were used. The reaction products were resolved by SDS–PAGE and visualized by fluorography. Point mutations in the N-terminal region of dTCF were introduced by site-directed mutagenesis in pGEX–dTCF(1–130). The same results were obtained after substitution with alanine (Fig. 4b) or glutamic acid (not shown). None of these substitutions affected binding to Armadillo. The K25A substitution was made in the same way in the UAS.dTCF transformation plasmid².

Received 2 June; accepted 13 July 1998.

1. Riese, J. et al. LEF-1, a nuclear factor coordinating signaling inputs from *wingless* and *decapentaplegic*. *Cell* **88**, 777–787 (1997).
2. van de Wetering, M. et al. Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. *Cell* **88**, 789–799 (1997).
3. Brunner, E., Peter, O., Schweizer, L. & Basler, K. *pangolin* encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature* **385**, 829–833 (1997).
4. Bienz, M. TCF: transcriptional activator or repressor? *Curr. Opin. Cell Biol.* **10**, 366–372 (1998).
5. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A β -catenin/Xtcf-3 complex binds to the *shimois* promoter to regulate dorsal axis specification in *Xenopus*. *Genes Dev.* **11**, 2359–2370 (1997).
6. Korinek, V. et al. Constitutive transcriptional activation by a β -catenin-Tcf complex in APC^{-/-} colon carcinoma. *Science* **275**, 1784–1787 (1997).

7. Morin, P. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutations in β -catenin or APC. *Science* **275**, 1787–1790 (1997).
8. Rubinfeld, B. *et al.* Stabilization of β -catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
9. Akimura, H. *et al.* *Drosophila* CBP is a co-activator of *cubitus interruptus* in *hedgehog* signalling. *Nature* **386**, 735–738 (1997).
10. Akimura, H., Hou, D.-X. & Ishii, S. *Drosophila* CBP is required for dorsal-dependent twist gene expression. *Nature Genet.* **17**, 211–214 (1997).
11. Chivria, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855–859 (1993).
12. Goldman, P. S., Tran, V. K. & Goodman, R. H. The multifunctional role of the co-activator CBP in transcriptional regulation. *Recent Prog. Horm. Res.* **52**, 103–119 (1997).
13. Bannister, A. J. & Kouzarides, T. CBP-induced stimulation of c-Fos activity is abrogated by E1A. *EMBO J.* **14**, 4758–4762 (1995).
14. Eresh, S., Riese, J., Jackson, D. B., Bohmann, D. & Bienz, M. A CREB binding site as a target for *decapentaplegic* signalling during *Drosophila* endoderm induction. *EMBO J.* **16**, 2014–2022 (1997).
15. Riese, J., Tremml, G. & Bienz, M. D-Fos, a target gene of *Decapentaplegic* signalling with a critical role during *Drosophila* endoderm induction. *Development* **124**, 3353–3361 (1997).
16. Perrimon, N. The genetic basis for patterned baldness in *Drosophila*. *Cell* **76**, 781–784 (1994).
17. Immerglück, K., Lawrence, P. A. & Bienz, M. Induction across germ layers in *Drosophila* mediated by a genetic cascade. *Cell* **62**, 261–268 (1990).
18. Hoppler, S. & Bienz, M. Two different thresholds of *wingless* signalling with distinct development consequences in the *Drosophila* midgut. *EMBO J.* **14**, 5016–5026 (1995).
19. Yu, X., Hoppler, S., Eresh, S. & Bienz, M. *decapentaplegic*, a target gene of the *wingless* signalling pathway in the *Drosophila* midgut. *Development* **122**, 849–858 (1996).
20. Siegfried, E., Chou, T. B. & Perrimon, N. *wingless* signaling acts through *zeste-white 3*, the *Drosophila* homolog of glycogen synthase kinase-3, to regulate *engrailed* and establish cell fate. *Cell* **71**, 1167–1179 (1992).
21. Szüts, D., Freeman, M. & Bienz, M. Antagonism between EGFR and *Wingless* signalling in the larval cuticle of *Drosophila*. *Development* **124**, 3209–3219 (1997).
22. Sanson, B., White, P. & Vincent, J.-P. Uncoupling cadherin-based adhesion from *wingless* signalling in *Drosophila*. *Nature* **383**, 627–630 (1996).
23. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
24. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
25. Imhof, A. *et al.* Acetylation of general transcription factors by histone acetyltransferases. *Curr. Biol.* **7**, 689–692 (1997).
26. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595–606 (1997).
27. Peifer, M., Sweeton, D., Casey, M. & Wieschaus, E. *wingless* signal and *Zeste-white 3* kinase trigger opposing changes in the intracellular distribution of Armadillo. *Development* **120**, 369–380 (1994).
28. Muraoka, M. *et al.* p300 gene alterations in colorectal and gastric carcinomas. *Oncogene* **12**, 1565–1569 (1996).
29. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank S. Ishii for *nej* mutants and plasmids; X. Yu, S. Greaves and J.-P. Vincent for strains and for sharing unpublished results; D. Owen for synthetic peptides; T. Kouzarides and R. Grosschedl for plasmids; M. Peifer and P. Simpson for fly strains; and X. Yu, M. Freeman and H. Pelham for discussion. L.W. is supported by an EMBO long-term fellowship.

Correspondence and requests for materials should be addressed to M.B. (e-mail: mb2@mrc-lmb.cam.ac.uk).

correction

Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis

Peter Carmeliet, Yuval Dor, Jean-Marc Herbert, Dai Fukumura, Koen Brusselmans, Mieke Dewerchin, Michal Neeman, Françoise Bono, Rinat Abramovitch, Patrick Maxwell, Cameron J. Koch, Peter Ratcliffe, Lieve Moons, Rakesh K. Jain, Désiré Collen & Eli Keshert

Nature **394**, 485–490 (1998)

The last author's name was misspelled but is now corrected above.

Also, in the penultimate sentence of the introductory bold paragraph, growth was accelerated in HIF-1 α ^{-/-} tumours (not in HIF-1 α tumours, as published). □

Tests of quantum gravity from observations of γ -ray bursts

G. Amelino-Camelia, John Ellis, N. E. Mavromatos, D. V. Nanopoulos & Subir Sarkar

Nature **393**, 763–765 (1998)

In ref. 22 of this Letter, the first author's name was incorrectly cited as C. L. Bhat instead of P. N. Bhat. □

erratum

Perinuclear localization of chromatin facilitates transcriptional silencing

Erik D. Andrulis, Aaron M. Neiman, David C. Zappulla & Rolf Sternglanz

Nature **394**, 592–595 (1998)

Owing to an error in the production process, the six bottom panels of Fig. 1 reproduced poorly. The figure is shown again here. □

